

*C*  
*CRSP-2*  
**NOVEL ~~CRSP~~ PROTEIN AND NUCLEIC ACID MOLECULES  
AND USES THEREFOR**

5

**Cross-Reference to Related Applications**

*INS C'*  
*INS C2*  
*C10*  
The present application ~~claims the benefit~~ of prior-filed application U.S. Patent Application Serial No. 08/842,898, entitled "Novel CRSP-1 Compositions and Therapeutic and Diagnostic Uses Therefor", filed April 17, 1997, and of prior-filed provisional application U.S. Patent Application No. 60/071,589, filed January 15, 1998 (Attorney Docket No.: MEI-008-1). The contents of each of the above-referenced patent applications are incorporated herein by this reference in their entirety.

**Background of the Invention**

15 Secreted proteins play an integral role in the formation, differentiation, and maintenance of cells in multicellular organisms. For instance, secretory proteins are known in the art to be involved in *signaling* between cells which are not in direct contact. Such secreted signaling molecules are particularly important in the development of vertebrate tissue during embryogenesis as well as in the maintenance of  
20 the differentiated state of adult tissues. For example, inductive interactions that occur between neighboring cell layers and tissues in the developing embryo are largely dependent on the existence and regulation of secreted signaling molecules. In inductive interactions, biochemical signals secreted by one cell population influence the developmental fate of a second cell population, typically by altering the fate of the  
25 second cell population. For example, the Wnt proteins are now recognized as one of the major families of developmentally important signaling molecules in organisms ranging from *Drosophila* to mice. For review see Cadigan, K.M. and R. Nusse (1997) *Genes & Development*, 11:3286-3305.

30 It is also now well recognized that major families of signaling molecules have a dual role to play in both the development of an organism as well as in promoting or maintaining the differentiated state of tissues in the adult animal. Furthermore, major families of signaling molecules have been implicated in controlling proliferation of cells in mature adult tissue, for example, during normal cell turnover in the adult organism as well as in tissue regeneration activated as a result of damage to the adult tissue.  
35

**Summary of the Invention**

The present invention is based, at least in part, on the discovery of nucleic acid molecules which encode a novel family of secreted proteins, referred to herein as the

Cysteine-Rich Secreted Proteins ("CRSPs" or "CRSP proteins"). The CRSP molecules of the present invention are useful as modulating agents in regulating a variety of cellular processes. Accordingly, in one aspect, this invention provides isolated nucleic acid molecules encoding CRSP proteins or biologically active portions thereof, as well as nucleic acid fragments suitable as primers or hybridization probes for the detection of CRSP-encoding nucleic acids.

In one embodiment, a CRSP nucleic acid molecule is 60% homologous to the nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:3, or complement thereof. In yet another embodiment, a CRSP nucleic acid molecule is 80% homologous to the nucleotide sequence shown in SEQ ID NO:4, SEQ ID NO:6, or a complement thereof. In yet another embodiment, a CRSP nucleic acid molecule is 60% homologous to the nucleotide sequence shown in SEQ ID NO:7, SEQ ID NO:9, the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number 98633, or a complement thereof. In yet another embodiment, a CRSP nucleic acid molecule is 85% homologous to the nucleotide sequence shown in SEQ ID NO:7, SEQ ID NO:9, the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number 98633, or a complement thereof. In yet another embodiment, a CRSP nucleic acid molecule is 70% homologous to the nucleotide sequence shown in SEQ ID NO:10, SEQ ID NO:12, or a complement thereof. In a preferred embodiment, an isolated CRSP nucleic acid molecule has the nucleotide sequence shown SEQ ID NO:3, or a complement thereof. In another embodiment, a CRSP nucleic acid molecule further comprises nucleotides 1-37 of SEQ ID NO:1. In yet another preferred embodiment, a CRSP nucleic acid molecule further comprises nucleotides 1088-2479 of SEQ ID NO:1. In another preferred embodiment, an isolated CRSP nucleic acid molecule has the nucleotide sequence shown in SEQ ID NO:1.

In another preferred embodiment, an isolated CRSP nucleic acid molecule has the nucleotide sequence shown SEQ ID NO:6, or a complement thereof. In another embodiment, a CRSP nucleic acid molecule further comprises nucleotides 1-125 of SEQ ID NO:4. In yet another preferred embodiment, a CRSP nucleic acid molecule further comprises nucleotides 797-848 of SEQ ID NO:4. In another preferred embodiment, an isolated CRSP nucleic acid molecule has the nucleotide sequence shown in SEQ ID NO:4.

In another preferred embodiment, an isolated CRSP nucleic acid molecule has the nucleotide sequence shown SEQ ID NO:9, the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number 98633, or a complement thereof. In another embodiment, a CRSP nucleic acid molecule further comprises nucleotides 1-92 of SEQ ID NO:7. In yet another preferred embodiment, a CRSP nucleic acid molecule further comprises nucleotides 891-1529 of SEQ ID NO:7.

In another preferred embodiment, an isolated CRSP nucleic acid molecule has the nucleotide sequence shown in SEQ ID NO:7.

5 In another preferred embodiment, an isolated CRSP nucleic acid molecule has the nucleotide sequence shown SEQ ID NO:12, or a complement thereof. In another embodiment, a CRSP nucleic acid molecule further comprises nucleotides 538-702 of SEQ ID NO:10. In yet another preferred embodiment, an isolated CRSP nucleic acid molecule has the nucleotide sequence shown in SEQ ID NO:10.

10 In another embodiment, a CRSP nucleic acid molecule includes a nucleotide sequence encoding a protein having an amino acid sequence sufficiently homologous to the amino acid sequence of SEQ ID NO:2, SEQ ID 5, SEQ ID NO:8, or SEQ ID NO:11. In another preferred embodiment, a CRSP nucleic acid molecule includes a nucleotide sequence encoding a protein having an amino acid sequence at least 60% homologous to the amino acid sequence of SEQ ID NO:2. In yet another preferred embodiment, a CRSP nucleic acid molecule includes a nucleotide sequence encoding a protein having  
15 an amino acid sequence at least 60% homologous to the amino acid sequence of SEQ ID NO:5. In yet another preferred embodiment, a CRSP nucleic acid molecule includes a nucleotide sequence encoding a protein having an amino acid sequence at least 60% homologous to the amino acid sequence of SEQ ID NO:8. In yet another preferred embodiment, a CRSP nucleic acid molecule includes a nucleotide sequence encoding a protein having an amino acid sequence at least 75% homologous to the amino acid sequence of SEQ ID NO:8. In yet another preferred embodiment, a CRSP nucleic acid molecule includes a nucleotide sequence encoding a protein having an amino acid sequence at least 65% homologous to the amino acid sequence of SEQ ID NO:11.

20 In another embodiment, an isolated nucleic acid molecule of the present invention encodes a CRSP protein which includes a signal sequence and a cysteine-rich region, and is secreted. In another embodiment, an isolated nucleic acid molecule of the present invention encodes a CRSP protein which includes a signal sequence and a cysteine-rich region, wherein the cysteine-rich region comprises at least one cysteine-rich domain, and is secreted. In yet another embodiment, a CRSP nucleic acid molecule  
25 encodes a CRSP protein and is a naturally occurring nucleotide sequence.

Another embodiment of the invention features CRSP nucleic acid molecules which specifically detect CRSP nucleic acid molecules relative to nucleic acid molecules encoding non-CRSP proteins. For example, in one embodiment, a CRSP nucleic acid molecule hybridizes under stringent conditions to a nucleic acid molecule comprising  
35 the nucleotides 470-2479 of nucleotide sequence shown in SEQ ID NO:1, to nucleotides 1-475 of nucleotide sequence shown in SEQ ID NO:4, or to nucleotides 1-600 of nucleotide sequence shown in SEQ ID NO:7, or hybridizes under stringent conditions to the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as

Accession Number \_\_\_\_\_, or to the nucleotide sequence of the DNA insert of the  
C plasmid deposited with ATCC as Accession Number 98633. In another embodiment,  
the CRSP nucleic acid molecule is at least 500 nucleotides in length and hybridizes  
under stringent conditions to a nucleic acid molecule comprising the nucleotide  
5 sequence shown in SEQ ID NO:1, SEQ ID NO:4, SEQ ID NO:7, or SEQ ID NO:10 or a  
complement thereof.

Another embodiment of the invention provides an isolated nucleic acid molecule  
which is antisense to the coding strand of a CRSP nucleic acid.

Another aspect of the invention provides a vector comprising a CRSP nucleic  
10 acid molecule. In certain embodiments, the vector is a recombinant expression vector.  
In another embodiment, the invention provides a host cell containing a vector of the  
invention. The invention also provides a method for producing a CRSP protein by  
culturing in a suitable medium, a host cell of the invention containing a recombinant  
expression vector such that a CRSP protein is produced.

15 Another aspect of this invention features isolated or recombinant CRSP proteins  
and polypeptides. In one embodiment, an isolated CRSP protein has a signal sequence  
and a cysteine-rich region which comprises two cysteine-rich domains, and is secreted.  
In another embodiment, an isolated CRSP protein has an amino acid sequence  
sufficiently homologous to the amino acid sequence of SEQ ID NO:2, SEQ ID NO:5,  
20 SEQ ID NO:8, or SEQ ID NO:11. In a preferred embodiment, a CRSP protein has an  
amino acid sequence at least about 60% homologous to the amino acid sequence of SEQ  
ID NO:2. In another preferred embodiment, a CRSP protein has an amino acid sequence  
at least about 60% homologous to the amino acid sequence of SEQ ID NO:5. In another  
preferred embodiment, a CRSP protein has an amino acid sequence at least about 60%  
25 homologous to the amino acid sequence of SEQ ID NO:8. In another preferred  
embodiment, a CRSP protein has an amino acid sequence at least about 75%  
homologous to the amino acid sequence of SEQ ID NO:8. In another preferred  
embodiment, a CRSP protein has an amino acid sequence at least about 65%  
homologous to the amino acid sequence of SEQ ID NO:11. In another embodiment, a  
30 CRSP protein has the amino acid sequence of SEQ ID NO:2, SEQ ID NO:5, SEQ ID  
NO:8, or SEQ ID NO:11.

Another embodiment of the invention features an isolated CRSP protein which is  
encoded by a nucleic acid molecule having a nucleotide sequence at least about 60%  
homologous to a nucleotide sequence of SEQ ID NO:1, or a complement thereof.

35 Another embodiment of the invention features an isolated CRSP protein which is  
encoded by a nucleic acid molecule having a nucleotide sequence at least about 80%  
homologous to a nucleotide sequence of SEQ ID NO:4, or a complement thereof.  
Another embodiment of the invention features an isolated CRSP protein which is

encoded by a nucleic acid molecule having a nucleotide sequence at least about 60% homologous to a nucleotide sequence of SEQ ID NO:7, or a complement thereof.

Another embodiment of the invention features an isolated CRSP protein which is encoded by a nucleic acid molecule having a nucleotide sequence at least about 85% homologous to a nucleotide sequence of SEQ ID NO:7, or a complement thereof.

Another embodiment of the invention features an isolated CRSP protein which is encoded by a nucleic acid molecule having a nucleotide sequence at least about 70% homologous to a nucleotide sequence of SEQ ID NO:10, or a complement thereof. This invention further features an isolated CRSP protein which is encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:4, SEQ ID NO:7, SEQ ID NO:10, or a complement thereof.

The CRSP proteins of the present invention, or biologically active portions thereof, can be operatively linked to a non-CRSP polypeptide to form CRSP fusion proteins. The invention further features antibodies that specifically bind CRSP proteins, such as monoclonal or polyclonal antibodies. In addition, the CRSP proteins or biologically active portions thereof can be incorporated into pharmaceutical compositions, which optionally include pharmaceutically acceptable carriers.

In another aspect, the present invention provides a method for detecting CRSP expression in a biological sample by contacting the biological sample with an agent capable of detecting a CRSP nucleic acid molecule, protein or polypeptide such that the presence of a CRSP nucleic acid molecule, protein or polypeptide is detected in the biological sample.

In another aspect, the present invention provides a method for detecting the presence of CRSP activity in a biological sample by contacting the biological sample with an agent capable of detecting an indicator of CRSP activity such that the presence of CRSP activity is detected in the biological sample.

In another aspect, the invention provides a method for modulating CRSP activity comprising contacting the cell with an agent that modulates CRSP activity such that CRSP activity in the cell is modulated. In one embodiment, the agent inhibits CRSP activity. In another embodiment, the agent stimulates CRSP activity. In one embodiment, the agent is an antibody that specifically binds to a CRSP protein. In another embodiment, the agent modulates expression of CRSP by modulating transcription of a CRSP gene or translation of a CRSP mRNA. In yet another embodiment, the agent is a nucleic acid molecule having a nucleotide sequence that is antisense to the coding strand of a CRSP mRNA or a CRSP gene.

In one embodiment, the methods of the present invention are used to treat a subject having a disorder characterized by aberrant CRSP protein or nucleic acid

expression or activity by administering an agent which is a CRSP modulator to the subject. In one embodiment, the CRSP modulator is a CRSP protein. In another embodiment the CRSP modulator is a CRSP nucleic acid molecule. In yet another embodiment, the CRSP modulator is a peptide, peptidomimetic, or other small molecule.

- 5 In a preferred embodiment, the disorder characterized by aberrant CRSP protein or nucleic acid expression is a developmental, differentiative, or proliferative disorder.

The present invention also provides a diagnostic assay for identifying the presence or absence of a genetic alteration characterized by at least one of (i) aberrant modification or mutation of a gene encoding a CRSP protein; (ii) mis-regulation of said  
10 gene; and (iii) aberrant post-translational modification of a CRSP protein, wherein a wild-type form of said gene encodes an protein with a CRSP activity.

In another aspect the invention provides a method for identifying a compound that binds to or modulates the activity of a CRSP protein, by providing a indicator composition comprising a CRSP protein having CRSP activity, contacting the indicator  
15 composition with a test compound, and determining the effect of the test compound on CRSP activity in the indicator composition to identify a compound that modulates the activity of a CRSP protein.

Other features and advantages of the invention will be apparent from the following detailed description and claims.

20

#### **Brief Description of the Drawings**

*Figure 1* depicts the cDNA sequence and predicted amino acid sequence of human CRSP-1. The nucleotide sequence corresponds to nucleic acids 1 to 2479 of SEQ ID NO:1. The amino acid sequence corresponds to amino acids 1 to 350 of SEQ  
25 ID NO:2.

*Figure 2* depicts the cDNA sequence and predicted amino acid sequence of human CRSP-2. The nucleotide sequence corresponds to nucleic acids 1 to 848 of SEQ ID NO:4. The amino acid sequence corresponds to amino acids 1 to 224 of SEQ ID NO:5.

30 *Figure 3* depicts the cDNA sequence and predicted amino acid sequence of human CRSP-3. The nucleotide sequence corresponds to nucleic acids 1 to 1529 of SEQ ID NO:7. The amino acid sequence corresponds to amino acids 1 to 266 of SEQ ID NO:8.

35 *Figure 4* depicts the cDNA sequence and predicted amino acid sequence of human CRSP-4. The nucleotide sequence corresponds to nucleic acids 1 to 702 of SEQ ID NO:10. The amino acid sequence corresponds to amino acids 1 to 179 of SEQ ID NO:11.

Figure 5 depicts the cDNA sequence and predicted amino acid sequence of human CRSP-like-N. The nucleotide sequence corresponds to nucleic acids 1 to 928 of SEQ ID NO:13. The amino acid sequence corresponds to amino acids 1 to 242 of SEQ ID NO:14.

5C Figure 6 is depicts the amino acid sequences of human CRSP-1, human CRSP-2, human CRSP-3, and human CRSP-4, as well as the consensus sequence generated from the alignment of these sequences. The figure details the relationship between the CRSP proteins of the instant invention.

10 C Figure 7 is a schematic diagram depicting the relationship between the CRSP proteins of the instant invention. The figure depicts the biological and functional domains of human CRSP. Signal peptides are indicated by filled boxes. The cysteine-rich domains of a CRSP cysteine-rich region are depicted as CRD-1 and CRD-2.

a Figure 8 depicts the cDNA sequence and predicted amino acid sequence of murine CRSP-1. The nucleotide sequence corresponds to nucleic acids 1 to 928 of SEQ ID NO:16. The amino acid sequence corresponds to amino acids 1 to 349 of SEQ ID NO:17.

### **Detailed Description of the Invention**

20 The present invention is based on the discovery of novel molecules, referred to herein as CRSP protein and nucleic acid molecules, which comprise a family of molecules having certain conserved structural and functional features. The term "family" when referring to the protein and nucleic acid molecules of the invention is intended to mean two or more proteins or nucleic acid molecules having a common

25 structural domain and having sufficient amino acid or nucleotide sequence homology as defined herein. Such family members can be naturally occurring and can be from either the same or different species. For example, a family can contain a first protein of human origin, as well as other, distinct proteins of human origin or alternatively, can contain homologues of non-human origin. Members of a family may also have common

30 functional characteristics.

In one embodiment, a CRSP family member is identified based on the presence of at least one "cysteine-rich region" in the protein or corresponding nucleic acid molecule. As used herein, the term "cysteine-rich region" refers to a protein domain having an amino acid sequence of about 120-200, preferably about 130-190, more

35 preferably about 140-180 amino acid residues, and even more preferably at least about 135-175 amino acids of which at least about 10-30, preferably about 15-20, and more preferably about 16, 17, 18, or 19 of the amino acids are cysteine residues. In a preferred embodiment, a cysteine-rich region is located in the C-terminal region of a

CRSP protein. For example, in one embodiment, a CRSP-1 protein contains a cysteine rich region containing about amino acids 147-284 of SEQ ID NO:2, having 20 cysteine residues at the positions indicated in Figure 6. In another embodiment, a CRSP-2 protein contains a cysteine rich region containing about amino acids 41-218 of SEQ ID NO:5, having 20 cysteine residues at the positions indicated in Figure 6. In another embodiment, a CRSP-3 protein contains a cysteine rich region containing about amino acids 85-263 of SEQ ID NO:8, having 20 cysteine residues at the positions indicated in Figure 6. In another embodiment, a CRSP-4 protein contains a cysteine rich region containing about amino acids 1-176 of SEQ ID NO:2, having 18 cysteine residues at the positions indicated in Figure 6.

In one embodiment, a cysteine-rich region of a CRSP protein of the present invention includes at least one cysteine-rich domain. As defined herein, a "cysteine-rich domain" refers to a protein domain having an amino acid sequence of about 30-100 amino acids, preferably about 35-95 amino acids, more preferably about 40-90 amino acids, more preferably about 45-85 amino acids, even more preferably about 50-80 amino acids, and even more preferably about 55-75, 60-70, or 65 amino acids, of which at least about 3-20, preferably about 5-15, or more preferably about 6, 7, 8, 9, 10, 11, or 12 amino acids are cysteine residues. A preferred CRSP protein of the present invention has two cysteine-rich-domains within a cysteine-rich region. For example, in one embodiment, a CRSP-1 protein contains a cysteine rich region including a first cysteine-rich domain containing about amino acids 147-195 of SEQ ID NO:2, having 10 cysteine residues, and a second cysteine-rich domain containing about amino acids 201-284 of SEQ ID NO:2, having 10 cysteine residues (the positions of the cysteine residues are depicted in Figure 6). In another embodiment, a CRSP-2 protein contains a cysteine rich region including a first cysteine-rich domain containing about amino acids 41-90 of SEQ ID NO:5, having 10 cysteine residues, and a second cysteine-rich domain containing about amino acids <sup>138-218</sup>~~41-218~~ of SEQ ID NO:5, having 10 cysteine residues (the positions of the cysteine residues are depicted in Figure 6). In another embodiment, a CRSP-3 protein contains a cysteine rich region including a first cysteine-rich domain containing about amino acids 85-138 of SEQ ID NO:8, having 10 cysteine residues, and a second cysteine-rich domain containing about amino acids 182-263 of SEQ ID NO:8, having 10 cysteine residues (the positions of the cysteine residues are depicted in Figure 6). In another embodiment, a CRSP-4 protein contains a cysteine rich region including a first cysteine-rich domain containing about amino acids 1-47 of SEQ ID NO:11, having 8 cysteine residues, and a second cysteine-rich domain containing about amino acids 96-176 of SEQ ID NO:11, having 10 cysteine residues (the positions of the cysteine residues are depicted in Figure 6).



In another embodiment, in addition to cysteine-rich domains, the cysteine-rich region contains a spacer region which separates the first and second cysteine-rich domains. As used herein, the "spacer region" refers to amino acid residues which are located between the first and second cysteine-rich domains of a cysteine-rich region and includes amino acid residues located C-terminal to the first cysteine-rich domain and N-terminal to the second cysteine-rich domain. As defined herein, a "spacer region" refers to a protein domain of about 5-70 amino acids, preferably about 10-65 amino acids, more preferably about 15-60 amino acids, even more preferably about 20-55 amino acids, and even more preferably about 25-50, 30-45 or 35-40 amino acids. For example, CRSP-1 protein contains a spacer region of about amino acids 196-200 of SEQ ID NO:2; CRSP-2 protein contains a spacer region of about amino acids 91-137 of SEQ ID NO:5; CRSP-3 protein contains a spacer region of about amino acids 139-181 of SEQ ID NO:8; and CRSP-4 protein contains a spacer region of about amino acids 48-95 of SEQ ID NO:11.

In another embodiment of the invention, the CRSP protein has at least one cysteine-rich region and a signal sequence. As used herein, a "signal sequence" refers to a peptide containing about 20 amino acids which occurs at the N-terminus of secretory and integral membrane proteins and which contains a large number of hydrophobic amino acid residues. For example, a signal sequence contains at least about 14-28 amino acid residues, preferably about 16-26 amino acid residues, more preferably about 18-24 amino acid residues, and more preferably about 20-22 amino acid residues, and has at least about 40-70%, preferably about 50-65%, and more preferably about 55-60% hydrophobic amino acid residues (e.g., Alanine, Valine, Leucine, Isoleucine, Phenylalanine, Tyrosine, Tryptophan, or Proline). Such a "signal sequence", also referred to in the art as a "signal peptide", serves to direct a protein containing such a sequence to a lipid bilayer. For example, in one embodiment, a CRSP-1 protein contains a signal sequence of about amino acids 1-23 of SEQ ID NO:2. In another embodiment, a CRSP-2 protein contains a signal sequence of about amino acids 1-19 of SEQ ID NO:5. In another embodiment, a CRSP-3 protein contains a signal sequence of about amino acids 1-20 of SEQ ID NO:8.

CRSP proteins of the present invention can be used to identify additional CRSP family members. For example, a protein having homology to CRSP-1 was identified using the nucleotide sequence encoding the N-terminal unique region of CRSP-1 to search a protein sequence database. Such a protein, referred to herein as "CRSP-like-N" is depicted in Figure 5. The nucleotide sequence of CRSP-like-N (SEQ ID NO:13) encodes a protein having 242 amino acids (SEQ ID NO:14). The nucleotide sequence of CRSP-like-N includes a 5' untranslated region containing nucleotides 1-74 of SEQ ID NO:13, a coding region containing <sup>nucleotides</sup> 75-800 of SEQ ID NO:13 (corresponding

to nucleotides 1-726 of SEQ ID NO:15), and a 3' untranslated region containing nucleotides 801-928 of SEQ ID NO:13.

Accordingly, one embodiment of the invention features a CRSP protein having at least one cysteine-rich region. Another embodiment features a CRSP protein having at least one cysteine-rich region, wherein the cysteine-rich region includes at least one cysteine-rich domain. Another embodiment features a CRSP protein having at least one cysteine-rich region, wherein the cysteine-rich region includes at least two cysteine-rich domains. Another embodiment features a protein having 20, 30, 40, 50, 60, 70, 80, 90, 95, or 99% homology to a cysteine-rich domain of a CRSP protein of the invention (e.g., CRSP-1, CRSP-2, CRSP-3, or CRSP-4).

Yet another embodiment of the invention features a CRSP protein having at least one cysteine-rich region and a signal peptide. Another embodiment features a CRSP protein having at least one cysteine-rich region and a signal peptide, wherein the cysteine-rich region includes at least one cysteine-rich domain. Another embodiment features a CRSP protein having at least one cysteine-rich region and a signal peptide, wherein the cysteine-rich region includes at least two cysteine-rich domains.

Preferred CRSP molecules of the present have an amino acid sequence sufficiently homologous to the amino acid sequence of SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8, or SEQ ID NO:11. As used herein, the term "sufficiently homologous" refers to a first amino acid or nucleotide sequence which contains a sufficient or minimum number of identical or equivalent (e.g., an amino acid residue which has a similar side chain) amino acid residues or nucleotides to a second amino acid or nucleotide sequence such that the first and second amino acid or nucleotide sequences share common structural domains and/or a common functional activity. For example, amino acid or nucleotide sequences which share common structural domains have at least about 40% homology, preferably 50% homology, more preferably 60%-70% homology across the amino acid sequences of the domains and contain at least one, preferably two, more preferably three, and even more preferably four, five or six structural domains, are defined herein as sufficiently homologous. Furthermore, amino acid or nucleotide sequences which share at least 40%, preferably 50%, more preferably 60, 70, or 80% homology and share a common functional activity are defined herein as sufficiently homologous.

As used interchangeably herein, a "CRSP activity", "biological activity of CRSP" or "functional activity of CRSP", refers to an activity exerted by a CRSP protein, polypeptide or nucleic acid molecule on a CRSP responsive cell as determined *in vivo*, or *in vitro*, according to standard techniques. In one embodiment, a CRSP activity is a direct activity, such as an association with a CRSP-target molecule. As used herein, a "target molecule" is a molecule with which a CRSP protein binds or interacts in nature,

- such that CRSP-mediated function is <sup>achieved</sup> ~~achieved~~. A CRSP target molecule can be a non-CRSP molecule or a CRSP protein or polypeptide of the present invention. In an exemplary embodiment, a CRSP target molecule is a membrane-bound protein (e.g., a "CRSP receptor") or a modified form of such a protein which has been altered such that the protein is soluble (e.g., recombinantly produced such that the protein does not express a membrane-binding domain). In another embodiment, a CRSP target is a second soluble protein molecule (e.g., a "CRSP binding partner" or "CRSP substrate"). In such an exemplary embodiment, a CRSP binding partner can be a second soluble non-CRSP protein or a second CRSP protein molecule of the present invention.
- Alternatively, a CRSP activity is an indirect activity, such as a cellular signaling activity mediated by interaction of the CRSP protein with a second protein (e.g., a CRSP receptor).

- In a preferred embodiment, a CRSP activity is at least one or more of the following activities: (i) interaction of a CRSP protein with a membrane-bound CRSP receptor; (ii) interaction of a CRSP protein with a soluble form of a CRSP receptor; (iii) interaction of a CRSP protein with an intracellular protein via a membrane-bound CRSP receptor; (iv) complex formation between a soluble CRSP protein and a second soluble CRSP binding partner; (iv) complex formation between a soluble CRSP protein and a second soluble CRSP binding partner, wherein the CRSP binding partner is a non-CRSP protein molecule; and (v) complex formation between a soluble CRSP protein and a second soluble CRSP binding partner, wherein the CRSP binding partner is a second CRSP protein molecule. In yet another preferred embodiment, a CRSP activity is at least one or more of the following activities: (1) modulation of cellular signal transduction, either *in vitro* or *in vivo*; (2) regulation of gene transcription in a cell involved in development or differentiation, either *in vitro* or *in vivo*; (3) regulation of gene transcription in a cell involved in ~~in~~ development or differentiation, wherein at least one gene encodes a differentiation-specific protein; (4) regulation of gene transcription in a cell involved in ~~in~~ development or <sup>differentiation</sup> differentiation, wherein at least one gene encodes a second secreted protein; (5) regulation of gene transcription in a cell involved in development or differentiation, wherein at least one gene encodes a signal transduction molecule; and (6) regulation of cellular proliferation, either *in vitro* or *in vivo*. As referred to herein, "differentiation-specific proteins" include proteins involved in the transition of a cell from the undifferentiated to the differentiated phenotype. For example, such proteins can be differentiation specific structural proteins or differentiation-specific transcription factors. Such differentiation-specific proteins are generally expressed at higher levels in cells which are making the transition from the undifferentiated to the differentiated phenotype (e.g., during embryonic development or during regeneration of mature tissue in the adult animal), or are <sup>expressed</sup> ~~expressed~~ at higher levels

in fully-differentiated or terminally-differentiated cells as compared to their undifferentiated counterparts. Also, as referred to herein, "differentiation-specific genes" include nucleic acid molecules which encode differentiation-specific proteins.

Accordingly, another embodiment of the invention features isolated CRSP  
5 proteins and polypeptides having a CRSP activity. Preferred CRSP proteins have at least one cysteine-rich region and a CRSP activity. In another preferred embodiment, the CRSP protein has at least one cysteine-rich region, wherein the cysteine-rich region comprises at least one cysteine-rich domain, and a CRSP activity. In another preferred  
10 embodiment, the CRSP protein has at least one cysteine-rich region, wherein the cysteine-rich region comprises at least two cysteine-rich domains, and a CRSP activity. In yet another preferred embodiment, a CRSP protein further comprises a signal sequence. In still another preferred embodiment, a CRSP protein has a cysteine-rich region, a CRSP activity, and an amino acid sequence sufficiently homologous to an amino acid sequence of SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8, or SEQ ID  
15 NO:11.

The human CRSP-1 cDNA, which is approximately 2479 nucleotides in length, encodes a protein which is approximately 350 amino acid residues in length. The human CRSP protein contains an N-terminal signal sequence and a cysteine-rich region comprising two cysteine-rich domains. A CRSP cysteine-rich region can be found at  
20 least, for example, from about amino acids 147-284 of SEQ ID NO:2. The CRSP-1 cysteine-rich region comprises a first cysteine-rich domain from about amino acids 147-195 of SEQ ID NO:2 and a second cysteine-rich domain from about amino acids 201-284 of SEQ ID NO:2. The human CRSP-1 protein is a secreted protein which further contains a signal sequence at about amino acids 1-23 of SEQ ID NO:2. The prediction  
25 of such a signal peptide can be made, for example, utilizing the computer algorithm SIGNALP (Henrik, et al. (1997) *Protein Engineering* 10:1-6).

The human CRSP-2 cDNA, which is approximately 848 nucleotides in length, encodes a protein which is approximately 224 amino acid residues in length. The human CRSP protein contains an N-terminal signal sequence and a cysteine-rich region  
30 comprising two cysteine-rich domains. A CRSP cysteine-rich region can be found at least, for example, from about amino acids 41-218 of SEQ ID NO:5. The CRSP-2 cysteine-rich region comprises a first cysteine-rich domain from about amino acids 41-90 of SEQ ID NO:5 and a second cysteine-rich domain from about amino acids 138-218 of SEQ ID NO:5. The human CRSP-2 protein is a secreted protein which further  
35 contains a signal sequence at about amino acids 1-19 of SEQ ID NO:5.

The human CRSP-3 cDNA, which is approximately 1529 nucleotides in length, encodes a protein which is approximately 266 amino acid residues in length. The human CRSP protein contains an N-terminal signal sequence and a cysteine-rich region

comprising two cysteine-rich domains. A CRSP cysteine-rich region can be found at least, for example, from about amino acids 85-263 of SEQ ID NO:8. The CRSP-3 cysteine-rich region comprises a first cysteine-rich domain from about amino acids 85-138 of SEQ ID NO:8 and a second cysteine-rich domain from about amino acids 182-263 of SEQ ID NO:2. The human CRSP-3 protein is a secreted protein which further contains a signal sequence at about amino acids 1-20 of SEQ ID NO:8.

The human CRSP-4 cDNA, which is approximately 702 nucleotides in length, encodes a protein which is approximately 179 amino acid residues in length. The human CRSP protein contains a cysteine-rich region comprising two cysteine-rich domains. A CRSP cysteine-rich region can be found at least, for example, from about amino acids 1-176 of SEQ ID NO:11. The CRSP-4 cysteine-rich region comprises a first cysteine-rich domain from about amino acids 1-47 of SEQ ID NO:11 and a second cysteine-rich domain from about amino acids 96-176 of SEQ ID NO:11.

Various aspects of the invention are described in further detail in the following subsections:

#### I. Isolated Nucleic Acid Molecules

One aspect of the invention pertains to isolated nucleic acid molecules that encode CRSP proteins or biologically active portions thereof, as well as nucleic acid fragments sufficient for use as hybridization probes to identify CRSP-encoding nucleic acids (e.g., CRSP mRNA) and fragments for use as PCR primers for the amplification or mutation of CRSP nucleic acid molecules. As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (e.g., cDNA or genomic DNA) and RNA molecules (e.g., mRNA) and analogs of the DNA or RNA generated using nucleotide analogs. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA.

An "isolated" nucleic acid molecule is one which is separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated CRSP nucleic acid molecule can contain less than about 5 kb, 4kb, 3kb, 2kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques,

or substantially free of chemical precursors or other chemicals when chemically synthesized.

A nucleic acid molecule of the present invention, e.g., a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:4, SEQ ID NO:7, or SEQ ID NO:10, the nucleotide sequence of the DNA insert of the plasmid deposited with  
5 ATCC as Accession Number 98633, the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number \_\_\_\_\_, or a portion thereof, can be isolated using standard molecular biology techniques and the sequence information provided herein. Using all or portion of the nucleic acid sequence of SEQ ID NO:1,  
10 SEQ ID NO:4, SEQ ID NO:7, or SEQ ID NO:10, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number \_\_\_\_\_, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as  
15 Accession Number 98633 as a hybridization probe, CRSP nucleic acid molecules can be isolated using standard hybridization and cloning techniques (e.g., as described in Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual*.  
2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).

Moreover, a nucleic acid molecule encompassing all or a portion of SEQ ID NO:1, SEQ ID NO:4, SEQ ID NO:7, or SEQ ID NO:10, or the nucleotide sequence of  
20 the DNA insert of the plasmid deposited with ATCC as Accession Number \_\_\_\_\_, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as  
Accession Number 98633 can be isolated by the polymerase chain reaction (PCR) using synthetic oligonucleotide primers designed based upon the sequence of SEQ ID NO:1, SEQ ID NO:4, SEQ ID NO:7, or SEQ ID NO:10, or the nucleotide sequence of the  
25 DNA insert of the plasmid deposited with ATCC as Accession Number \_\_\_\_\_, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as  
Accession Number 98633.

A nucleic acid of the invention can be amplified using cDNA, mRNA or alternatively, genomic DNA, as a template and appropriate oligonucleotide primers  
30 according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to CRSP nucleotide sequences can be prepared by standard synthetic techniques, e.g., using an automated DNA synthesizer.

In a preferred embodiment, an isolated nucleic acid molecule of the invention  
35 comprises the nucleotide sequence shown in SEQ ID NO:1. The sequence of SEQ ID NO:1 corresponds to the human CRSP-1 cDNA. This cDNA comprises sequences encoding the human CRSP-1 protein (i.e., "the coding region", from nucleotides 38-1087), as well as 5' untranslated sequences (nucleotides 1 to 37) and 3' untranslated

sequences (nucleotides 1088 to 2479). Alternatively, the nucleic acid molecule can comprise only the coding region of SEQ ID NO:1 (e.g., nucleotides 38 to 1087, corresponding to SEQ ID NO:3).

5 In another preferred embodiment, an isolated nucleic acid molecule of the invention comprises the nucleotide sequence shown in SEQ ID NO:4. The sequence of SEQ ID NO:4 corresponds to the human CRSP-2 cDNA. This cDNA comprises sequences encoding the human CRSP-2 protein (i.e., "the coding region", from nucleotides 126-796), as well as 5' untranslated sequences (nucleotides 1 to 125) and 3' untranslated sequences (nucleotides 797 to 848). Alternatively, the nucleic acid  
10 molecule can comprise only the coding region of SEQ ID NO:4 (e.g., nucleotides 126 to 796, corresponding to SEQ ID NO:6).

In another preferred embodiment, an isolated nucleic acid molecule of the invention comprises the nucleotide sequence shown in SEQ ID NO:7. The sequence of SEQ ID NO:7 corresponds to the human CRSP-3 cDNA. This cDNA comprises  
15 sequences encoding the human CRSP-3 protein (i.e., "the coding region", from nucleotides 93-890), as well as 5' untranslated sequences (nucleotides 1 to 92) and 3' untranslated sequences (nucleotides 891 to 1529). Alternatively, the nucleic acid molecule can comprise only the coding region of SEQ ID NO:7 (e.g., nucleotides 93 to 890, corresponding to SEQ ID NO:9).

20 In another preferred embodiment, an isolated nucleic acid molecule of the invention comprises the nucleotide sequence shown in SEQ ID NO:10. The sequence of SEQ ID NO:10 corresponds to the human CRSP-4 cDNA. This cDNA comprises sequences encoding the human CRSP-4 protein (i.e., "the coding region", from nucleotides 1-537), as well as 3' untranslated sequences (nucleotides 538 to 702).  
25 Alternatively, the nucleic acid molecule can comprise only the coding region of SEQ ID NO:10 (e.g., nucleotides 1 to 537, corresponding to SEQ ID NO:12).

In another preferred embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule which is a complement of the nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:4, SEQ ID NO:7, or SEQ ID NO:10, the  
30 nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number \_\_\_\_\_, the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number 98633, or a portion of any of these nucleotide sequences. A nucleic acid molecule which is complementary to the nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:4, SEQ ID NO:7, or SEQ ID  
35 NO:10, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number \_\_\_\_\_, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number 98633, is one which is sufficiently complementary to the nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:4, SEQ

ID NO:7, or SEQ ID NO:10, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number \_\_\_\_\_, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number 98633, such that it can hybridize to the nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:4, SEQ ID NO:7, or SEQ ID NO:10, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number \_\_\_\_\_, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number 98633, thereby forming a stable duplex.

In still another preferred embodiment, an isolated nucleic acid molecule of the present invention comprises a nucleotide sequence which is at least about 60-65%, preferably at least about 70-75%, more preferable at least about 80-85%, and even more preferably at least about 90-95% or more homologous to the nucleotide sequences shown in SEQ ID NO:1, SEQ ID NO:4, SEQ ID NO:7, or SEQ ID NO:10, the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number \_\_\_\_\_, the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number 98633, or a portion of any of these nucleotide sequences.

Moreover, the nucleic acid molecule of the invention can comprise only a portion of the nucleic acid sequence of SEQ ID NO:1, SEQ ID NO:4, SEQ ID NO:7, SEQ ID NO:10, the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number \_\_\_\_\_, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number 98633, for example a fragment which can be used as a probe or primer or a fragment encoding a biologically active portion of a CRSP protein. The nucleotide sequence determined from the cloning of the human CRSP genes allows for the generation of probes and primers designed for use in identifying and/or cloning CRSP homologues in other cell types, e.g., from other tissues, as well as CRSP homologues from other mammals. The probe/primer typically comprises substantially purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, preferably about 25, more preferably about 40, 50 or 75 consecutive nucleotides of a sense sequence of SEQ ID NO:1, SEQ ID NO:4, SEQ ID NO:7, SEQ ID NO:10, the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number \_\_\_\_\_, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number 98633, of an anti-sense sequence of SEQ ID NO:1, SEQ ID NO:4, SEQ ID NO:7, SEQ ID NO:10, the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number \_\_\_\_\_, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number 98633, or of a naturally occurring mutant of SEQ ID



NO:1, SEQ ID NO:4, SEQ ID NO:7, SEQ ID NO:10, the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number \_\_\_\_\_, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number 98633. In an exemplary embodiment, a nucleic acid molecule of the present invention comprises a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid molecule comprising nucleotides 470-2479 of SEQ ID NO:1 or to a nucleic acid molecule comprising nucleotides 1-475 of SEQ ID NO:4.

Probes based on the human CRSP nucleotide sequence can be used to detect transcripts or genomic sequences encoding the same or homologous proteins. In preferred embodiments, the probe further comprises a label group attached thereto, e.g., the label group can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as a part of a diagnostic test kit for identifying cells or tissue which misexpress a CRSP protein, such as by measuring a level of a CRSP-encoding nucleic acid in a sample of cells from a subject e.g., detecting CRSP mRNA levels or determining whether a genomic CRSP gene has been mutated or deleted.

A nucleic acid fragment encoding a "biologically active portion of a CRSP protein" can be prepared by isolating a portion of SEQ ID NO:1, SEQ ID NO:4, SEQ ID NO:7, SEQ ID NO:10, the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number \_\_\_\_\_, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number 98633, which encodes a polypeptide having a CRSP biological activity (the biological activities of the CRSP proteins have previously been described), expressing the encoded portion of the CRSP protein (e.g., by recombinant expression *in vitro*) and assessing the activity of the encoded portion of the CRSP protein.

The invention further encompasses nucleic acid molecules that differ from the nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:4, SEQ ID NO:7, SEQ ID NO:10, the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number \_\_\_\_\_, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number 98633 due to degeneracy of the genetic code and thus encode the same CRSP proteins as those encoded by the nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:4, SEQ ID NO:7, SEQ ID NO:10, the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number \_\_\_\_\_, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number 98633. In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence shown in SEQ ID NO:2, SEQ ID NO: 5, SEQ ID NO:8, or SEQ ID NO:11.

In addition to the human CRSP nucleotide sequences shown in SEQ ID NO:1, SEQ ID NO:4, SEQ ID NO:7, SEQ ID NO:10, the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number \_\_\_\_\_, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number 98633, it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of the CRSP proteins may exist within a population (e.g., the human population). Such genetic polymorphism in the CRSP genes may exist among individuals within a population due to natural allelic variation. As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules comprising an open reading frame encoding a CRSP protein, preferably a mammalian CRSP protein. Such natural allelic variations can typically result in 1-5% variance in the nucleotide sequence of a CRSP gene. Any and all such nucleotide variations and resulting amino acid polymorphisms in CRSP genes that are the result of natural allelic variation and that do not alter the functional activity of a CRSP protein are intended to be within the scope of the invention.

Moreover, nucleic acid molecules encoding CRSP proteins from other species, and thus which have a nucleotide sequence which differs from the human sequence of SEQ ID NO:1, SEQ ID NO:4, SEQ ID NO:7, SEQ ID NO:10, the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number \_\_\_\_\_, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number 98633 are intended to be within the scope of the invention. For example, a murine CRSP-1 cDNA has been identified based of the nucleotide sequence of human CRSP-1. The nucleotide sequence of murine CRSP-1 (SEQ ID NO:16) encodes a CRSP-1 protein having 349 amino acids. The nucleotide and amino acid sequences of murine CRSP-1 are depicted in Figure 8. The coding region of murine CRSP-1 is represented by SEQ ID NO:18.

Nucleic acid molecules corresponding to natural allelic variants and homologues of the CRSP cDNAs of the invention can be isolated based on their homology to the human CRSP nucleic acids disclosed herein using the human cDNA, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions.

Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention is at least 15 nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:4, SEQ ID NO:7, SEQ ID NO:10, the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number \_\_\_\_\_, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number 98633. In other embodiment, the nucleic acid is at least 30, 50, 100, 250 or 500 nucleotides in

length. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 60% homologous to each other typically remain hybridized to each other.

Preferably, the conditions are such that sequences at least about 70%, more preferably at least about 80%, even more preferably at least about 85% or 90% homologous to each other typically remain hybridized to each other. Such stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. A preferred, non-limiting example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2 X SSC, 0.1% SDS at 50-65°C. Preferably, an isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to the sequence of SEQ ID NO:1 corresponds to a naturally-occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (e.g., encodes a natural protein).

In addition to naturally-occurring allelic variants of the CRSP sequences that may exist in the population, the skilled artisan will further appreciate that changes can be introduced by mutation into the nucleotide sequences of SEQ ID NO:1, SEQ ID NO:4, SEQ ID NO:7, SEQ ID NO:10, the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number \_\_\_\_\_, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number 98633, thereby leading to changes in the amino acid sequence of the encoded CRSP proteins, without altering the functional ability of the CRSP proteins. For example, nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues can be made in the sequence of SEQ ID NO:1, SEQ ID NO:4, SEQ ID NO:7, SEQ ID NO:10, the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number \_\_\_\_\_, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number 99633. A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequence of CRSP (e.g., the sequence of SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8, or SEQ ID NO:11) without altering the biological activity, whereas an "essential" amino acid residue is required for biological activity. For example, amino acid residues that are conserved among the CRSP proteins of the present invention, are predicted to be particularly unamenable to alteration. Furthermore, amino acid residues that are conserved between CRSP protein and other proteins having cysteine-rich domains are not likely to be amenable to alteration.

Accordingly, another aspect of the invention pertains to nucleic acid molecules encoding CRSP proteins that contain changes in amino acid residues that are not

essential for activity. Such CRSP proteins differ in amino acid sequence from SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8, or SEQ ID NO:11 yet retain biological activity. In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least about  
5 60% homologous to the amino acid sequence of SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8, or SEQ ID NO:11. Preferably, the protein encoded by the nucleic acid molecule is at least about 65-70% homologous to SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8, or SEQ ID NO:11, more preferably at least about 75-80% homologous to SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8, or SEQ ID NO:11, even more preferably at least about  
10 85-90% homologous to SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8, or SEQ ID NO:11, and most preferably at least about 95% homologous to SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8, or SEQ ID NO:11.

An isolated nucleic acid molecule encoding a CRSP protein homologous to the protein of SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8, or SEQ ID NO:11 can be  
15 created by introducing one or more nucleotide substitutions, additions or deletions into the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:4, SEQ ID NO:7, SEQ ID NO:10, the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number \_\_\_\_\_, or the nucleotide sequence of the DNA insert of the  
B plasmid deposited with ATCC as Accession Number 98633, such that one or more  
20 amino acid substitutions, additions or deletions are introduced into the encoded protein. Mutations can be introduced into SEQ ID NO:1, SEQ ID NO:4, SEQ ID NO:7, SEQ ID NO:10, the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number \_\_\_\_\_, or the nucleotide sequence of the DNA insert of the  
C plasmid deposited with ATCC as Accession Number 98633 by standard techniques,  
25 such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art.  
30 These families include amino acids with basic side chains (e.g., lysine, arginine, histidine), acidic side chains (e.g., aspartic acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (e.g., threonine, valine, isoleucine)  
35 and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted nonessential amino acid residue in a CRSP protein is preferably replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of a CRSP coding

sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for CRSP biological activity to identify mutants that retain activity. Following mutagenesis of SEQ ID NO:1, SEQ ID NO:4, SEQ ID NO:7, SEQ ID NO:10, the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as  
 5 Accession Number \_\_\_\_\_, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number 98633, the encoded protein can be expressed recombinantly and the activity of the protein can be determined.

In a preferred embodiment, a mutant CRSP protein can be assayed for (1) the ability to modulate cellular signal transduction, either *in vitro* or *in vivo*; (2) regulate  
 10 gene transcription in a cell involved in development or differentiation, either *in vitro* or *in vivo*; (3) regulate gene transcription in a cell involved in ~~in~~ development or differentiation, wherein at least one gene encodes a differentiation-specific protein; (4)  
 15 regulate gene transcription in a cell involved in in development or ~~differentiation~~, wherein at least one gene encodes a second secreted protein; (5) regulate gene transcription in a cell involved in development or differentiation, wherein at least one gene encodes a signal transduction molecule; and (6) regulate cellular proliferation, either *in vitro* or *in vivo*.

In addition to the nucleic acid molecules encoding CRSP proteins described above, another aspect of the invention pertains to isolated nucleic acid molecules which  
 20 are antisense thereto. An "antisense" nucleic acid comprises a nucleotide sequence which is complementary to a "sense" nucleic acid encoding a protein, e.g., complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. Accordingly, an antisense nucleic acid can hydrogen bond to a sense nucleic acid. The antisense nucleic acid can be  
 25 complementary to an entire CRSP coding strand, or to only a portion thereof. In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding CRSP. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues (e.g., the coding region of human CRSP-1 corresponds to SEQ  
 30 ID NO:3, the coding region of human CRSP-2 corresponds to SEQ ID NO:6, the coding region of human CRSP-3 corresponds to SEQ ID NO:9, the coding region of human CRSP-4 corresponds to SEQ ID NO:12, and the coding region of human CRSP-like-N corresponds to SEQ ID NO:15). In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide  
 35 sequence encoding CRSP. The term "noncoding region" refers to 5' and 3' sequences which flank the coding region that are not translated into amino acids (i.e., also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding CRSP disclosed herein (e.g., SEQ ID NO:3, SEQ ID NO:6, SEQ ID NO:9, or SEQ ID NO:12), antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of CRSP mRNA, but more preferably is an oligonucleotide which is antisense to only a portion of the coding or noncoding region of CRSP mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of CRSP mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (e.g., an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, e.g., phosphorothioate derivatives and acridine substituted nucleotides can be used. Examples of modified nucleotides which can be used to generate the antisense nucleic acid include 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xantine, 4-acetylcytosine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (i.e., RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a CRSP protein to thereby inhibit expression of the protein, e.g., by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the

case of an antisense nucleic acid molecule which binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention include direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface, e.g., by linking the antisense nucleic acid molecules to peptides or antibodies which bind to cell surface receptors or antigens. The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of the antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an  $\alpha$ -anomeric nucleic acid molecule. An  $\alpha$ -anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual  $\beta$ -units, the strands run parallel to each other (Gaultier et al. (1987) *Nucleic Acids Res.* 15:6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue et al. (1987) *Nucleic Acids Res.* 15:6131-6148) or a chimeric RNA-DNA analogue (Inoue et al. (1987) *FEBS Lett.* 215:327-330).

In still another embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity which are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (e.g., hammerhead ribozymes (described in Haselhoff and Gerlach (1988) *Nature* 334:585-591)) can be used to catalytically cleave CRSP mRNA transcripts to thereby inhibit translation of CRSP mRNA. A ribozyme having specificity for a CRSP-encoding nucleic acid can be designed based upon the nucleotide sequence of a CRSP cDNA disclosed herein (i.e., SEQ ID NO:1, SEQ ID NO:4, SEQ ID NO:7, SEQ ID NO:10, the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number \_\_\_\_\_, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number 98033). For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a CRSP-encoding mRNA. See, e.g., Cech et al. U.S. Patent No. 4,987,071; and Cech et al. U.S. Patent No. 5,116,742. Alternatively, CRSP mRNA can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, e.g., Bartel, D. and Szostak, J.W. (1993) *Science* 261:1411-1418.

Alternatively, CRSP gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the CRSP (e.g., the CRSP promoter and/or enhancers) to form triple helical structures that prevent transcription of the CRSP gene in target cells. See generally, Helene, C. (1991) *Anticancer Drug Des.* 6(6):569-84; Helene, C. et al. (1992) *Ann. N.Y. Acad. Sci.* 660:27-36; and Maher, L.J. (1992) *Bioassays* 14(12):807-15.

In yet another embodiment, the CRSP nucleic acid molecules of the present invention can be modified at the base moiety, sugar moiety or phosphate backbone to improve, e.g., the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acid molecules can be modified to generate peptide nucleic acids (see Hyrup B. et al. (1996) *Bioorganic & Medicinal Chemistry* 4 (1): 5-23). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics, e.g., DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup B. et al. (1996) *supra*; Perry-O'Keefe et al. *PNAS* 93: 14670-675.

PNAs of CRSP nucleic acid molecules can be used therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, for example, inducing transcription or translation arrest or inhibiting replication. PNAs of CRSP nucleic acid molecules can also be used in the analysis of single base pair mutations in a gene, (e.g., by PNA-directed PCR clamping); as 'artificial restriction enzymes' when used in combination with other enzymes, (e.g., S1 nucleases (Hyrup B. (1996) *supra*)); or as probes or primers for DNA sequencing or hybridization (Hyrup B. et al. (1996) *supra*; Perry-O'Keefe *supra*).

In another embodiment, PNAs of CRSP can be modified, (e.g., to enhance their stability or cellular uptake), by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras of CRSP nucleic acid molecules can be generated which may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes, (e.g., RNase H and DNA polymerases), to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (Hyrup B. (1996) *supra*). The synthesis of PNA-DNA



chimeras can be performed as described in Hyrup B. (1996) *supra* and Finn P.J. *et al.* (1996) *Nucleic Acids Res.* 24 (17): 3357-63. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry and modified nucleoside analogs, e.g., 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine  
5 phosphoramidite, can be used as a between the PNA and the 5' end of DNA (Mag, M. *et al.* (1989) *Nucleic Acid Res.* 17: 5973-88). PNA monomers are then coupled in a stepwise manner to produce a chimeric <sup>molecule</sup> molecule with a 5' PNA segment and a 3' DNA segment (Finn P.J. *et al.* (1996) *supra*). Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment (Peterser, K.H. *et al.* (1975)  
10 *Bioorganic Med. Chem. Lett.* 5: 1119-11124).

In other embodiments, the oligonucleotide may include other appended groups such as peptides (e.g., for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (see, e.g., Letsinger *et al.* (1989) *Proc. Natl. Acad. Sci. US.* 86:6553-6556; Lemaitre *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:648-652;  
15 PCT Publication No. W088/09810, published December 15, 1988) or the blood-brain barrier (see, e.g., PCT Publication No. W089/10134, published April 25, 1988). In addition, oligonucleotides can be modified with hybridization-triggered cleavage agents (See, e.g., Krol *et al.* (1988) *BioTechniques* 6:958-976) or intercalating agents. (See, e.g., Zon (1988) *Pharm. Res.* 5:539-549). To this end, the oligonucleotide may be  
20 conjugated to another molecule, (e.g., a peptide, hybridization triggered cross-linking agent, transport agent, or hybridization-triggered cleavage agent).

## II. Isolated CRSP Proteins and Anti-CRSP Antibodies

One aspect of the invention pertains to isolated CRSP proteins, and biologically  
25 active portions thereof, as well as polypeptide fragments suitable for use as immunogens to raise anti-CRSP antibodies. In one embodiment, native CRSP proteins can be isolated from cells or tissue sources by an appropriate purification scheme using standard protein purification techniques. In another embodiment, CRSP proteins are produced by recombinant DNA techniques. Alternative to recombinant expression, a  
30 CRSP protein or polypeptide can be synthesized chemically using standard peptide synthesis techniques.

An "isolated" or "purified" protein or biologically active portion thereof is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the CRSP protein is derived, or substantially free from  
35 chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of CRSP protein in which the protein is separated from cellular components of the cells from which it is isolated or recombinantly produced. In one embodiment, the language "substantially free of

cellular material" includes preparations of CRSP protein having less than about 30% (by dry weight) of non-CRSP protein (also referred to herein as a "contaminating protein"), more preferably less than about 20% of non-CRSP protein, still more preferably less than about 10% of non-CRSP protein, and most preferably less than about 5% non-CRSP protein. When the CRSP protein or biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, i.e., culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the protein preparation.

The language "substantially free of chemical precursors or other chemicals" includes preparations of CRSP protein in which the protein is separated from chemical precursors or other chemicals which are involved in the synthesis of the protein. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of CRSP protein having less than about 30% (by dry weight) of chemical precursors or non-CRSP chemicals, more preferably less than about 20% chemical precursors or non-CRSP chemicals, still more preferably less than about 10% chemical precursors or non-CRSP chemicals, and most preferably less than about 5% chemical precursors or non-CRSP chemicals.

Biologically active portions of a CRSP protein include peptides comprising amino acid sequences sufficiently homologous to or derived from the amino acid sequence of the CRSP protein, e.g., the amino acid sequence shown in SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8, or SEQ ID NO:11, which include less amino acids than the full length CRSP proteins, and exhibit at least one activity of a CRSP protein. Typically, biologically active portions comprise a domain or motif with at least one activity of the CRSP protein. A biologically active portion of a CRSP protein can be a polypeptide which is, for example, 10, 25, 50, 100 or more amino acids in length.

In one embodiment, a biologically active portion of a CRSP protein comprises at least a cysteine-rich region. In another embodiment, a biologically active portion of a CRSP protein comprises at least a cysteine-rich region, wherein the cysteine-rich region includes at least one cysteine-rich domain. In yet another embodiment, a biologically active portion of a CRSP protein comprises at least a signal sequence.

In an alternative embodiment, a biologically active portion of a CRSP protein comprises a CRSP amino acid sequence lacking a signal sequence.

It is to be understood that a preferred biologically active portion of a CRSP protein of the present invention may contain at least one of the above-identified structural domains. A more preferred biologically active portion of a CRSP protein may contain at least two of the above-identified structural domains. An even more preferred biologically active portion of a CRSP protein may contain at least three of the above-identified structural domains. A particularly preferred biologically active portion

of a CRSP protein of the present invention may contain at least four of the above-identified structural domains.

Moreover, other biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of a native CRSP protein.

In a preferred embodiment, the CRSP protein has an amino acid sequence shown in SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8, or SEQ ID NO:11. In other embodiments, the CRSP protein is substantially homologous to SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8, or SEQ ID NO:11, and retains the functional activity of the protein of SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8, or SEQ ID NO:11, yet differs in amino acid sequence due to natural allelic variation or mutagenesis, as described in detail in subsection I above. Accordingly, in another embodiment, the CRSP protein is a protein which comprises an amino acid sequence at least about 60% homologous to the amino acid sequence of SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8, or SEQ ID NO:11 and retains the functional activity of the CRSP proteins of SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8, or SEQ ID NO:11, respectively. Preferably, the protein is at least about 70% homologous to SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8, or SEQ ID NO:11, more preferably at least about 80% homologous to SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8, or SEQ ID NO:11, even more preferably at least about 90% homologous to SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8, or SEQ ID NO:11, and most preferably at least about 95% or more homologous to SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8, or SEQ ID NO:11.

To determine the percent homology of two amino acid sequences or of two nucleic acids, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in the sequence of a first amino acid or nucleic acid sequence for optimal alignment with a second amino or nucleic acid sequence). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are homologous at that position (i.e., as used herein amino acid or nucleic acid "homology" is equivalent to amino acid or nucleic acid "identity"). The percent homology between the two sequences is a function of the number of identical positions shared by the sequences (i.e.,  $\% \text{ homology} = \# \text{ of identical positions} / \text{total } \# \text{ of positions} \times 100$ ). The determination of percent homology between two sequences can be accomplished using a mathematical algorithm. A preferred, non-limiting example of a mathematical algorithm utilized for the comparison of two sequences is the algorithm of Karlin and Altschul (1990) Proc. Natl. Acad. Sci. USA 87:2264-68, modified as in Karlin and Altschul (1993) Proc. Natl. Acad. Sci. USA 90:5873-77. Such an algorithm

is incorporated into the NBLAST and XBLAST programs of Altschul, et al. (1990) J. Mol. Biol. 215:403-10. BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to CRSP nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to CRSP protein molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul et al., (1997) Nucleic Acids Research 25(17):3389-3402. When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used. See <http://www.ncbi.nlm.nih.gov>.

The invention also provides CRSP chimeric or fusion proteins. As used herein, a CRSP "chimeric protein" or "fusion protein" comprises a CRSP polypeptide operatively linked to a non-CRSP polypeptide. A "CRSP polypeptide" refers to a polypeptide having an amino acid sequence corresponding to CRSP, whereas a "non-CRSP polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein which is not substantially homologous to the CRSP protein, e.g., a protein which is different from the CRSP protein and which is derived from the same or a different organism. Within a CRSP fusion protein the CRSP polypeptide can correspond to all or a portion of a CRSP protein. In a preferred embodiment, a CRSP fusion protein comprises at least one biologically active portion of a CRSP protein. In another preferred embodiment, a CRSP fusion protein comprises at least two biologically active portions of a CRSP protein. In another preferred embodiment, a CRSP fusion protein comprises at least three biologically active portions of a CRSP protein. Within the fusion protein, the term "operatively linked" is intended to indicate that the CRSP polypeptide and the non-CRSP polypeptide are fused in-frame to each other. The non-CRSP polypeptide can be fused to the N-terminus or C-terminus of the CRSP polypeptide.

For example, in one embodiment, the fusion protein is a GST-CRSP fusion protein in which the CRSP sequences are fused to the C-terminus of the GST sequences. Such fusion proteins can facilitate the purification of recombinant CRSP.

In another embodiment, the fusion protein is a CRSP protein containing a heterologous signal sequence at its N-terminus. For example, the native CRSP signal sequence (i.e., about amino acids 1 to 23 of SEQ ID NO:2) can be removed and replaced with a signal sequence from another protein. In certain host cells (e.g., mammalian host cells), expression and/or secretion of CRSP can be increased through use of a heterologous signal sequence.

In yet another embodiment, the fusion protein is a CRSP-immunoglobulin fusion protein in which the CRSP sequences comprising primarily the CRSP cysteine-rich

regions are fused to sequences derived from a member of the immunoglobulin protein family. Soluble derivatives have also been made of cell surface glycoproteins in the immunoglobulin gene superfamily consisting of an extracellular domain of the cell surface glycoprotein fused to an immunoglobulin constant (Fc) region (see e.g., Capon, D.J. *et al.* (1989) *Nature* 337:525-531 and Capon U.S. Patents 5,116,964 and 5,428,130 [CD4-IgG1 constructs]; Linsley, P.S. *et al.* (1991) *J. Exp. Med.* 173:721-730 [a CD28-IgG1 construct and a B7-1-IgG1 construct]; and Linsley, P.S. *et al.* (1991) *J. Exp. Med.* 174:561-569 and U.S. Patent 5,434,131[a CTLA4-IgG1]). Such fusion proteins have proven useful for modulating receptor-ligand interactions. Soluble derivatives of cell surface proteins of the tumor necrosis factor receptor (TNFR) superfamily proteins have been made consisting of an extracellular domain of the cell surface receptor fused to an immunoglobulin constant (Fc) region (See for example Moreland *et al.* (1997) *N. Engl. J. Med.* 337(3):141-147; van der Poll *et al.* (1997) *Blood* 89(10):3727-3734; and Ammann *et al.* (1997) *J. Clin. Invest.* 99(7):1699-1703.)

The CRSP-immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject to inhibit an interaction between a CRSP ligand and a CRSP receptor on the surface of a cell, to thereby suppress CRSP-mediated signal transduction *in vivo*. The CRSP-immunoglobulin fusion proteins can be used to affect the bioavailability of a CRSP cognate receptor. Inhibition of the CRSP ligand/CRSP interaction may be useful therapeutically for both the treatment of differentiative or proliferative disorders, as well as modulating (e.g., promoting or inhibiting) developmental responses, cell adhesion, and/or cell fate. Moreover, the CRSP-immunoglobulin fusion proteins of the invention can be used as immunogens to produce anti-CRSP antibodies in a subject, to purify CRSP ligands and in screening assays to identify molecules which inhibit the interaction of CRSP with a CRSP ligand.

Preferably, a CRSP chimeric or fusion protein of the invention is produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, for example by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and reamplified to generate a chimeric gene sequence (see, for example, *Current Protocols in Molecular Biology*, eds. Ausubel

et al. John Wiley & Sons: 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g., a GST polypeptide). A CRSP-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the CRSP protein.

5       The present invention also pertains to variants of the CRSP proteins which function as either CRSP agonists (mimetics) or as CRSP antagonists. Variants of the CRSP proteins can be generated by mutagenesis, e.g., discrete point mutation or truncation of a CRSP protein. An agonist of the CRSP proteins can retain substantially the same, or a subset, of the biological activities of the naturally occurring form of a  
10   CRSP protein. An antagonist of a CRSP protein can inhibit one or more of the activities of the naturally occurring form of the CRSP protein by, for example, competitively binding to a downstream or upstream member of a cellular signaling cascade which includes the CRSP protein. Thus, specific biological effects can be elicited by treatment  
15   with a variant of limited function. In one embodiment, treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein has fewer side effects in a subject relative to treatment with the naturally occurring form of the CRSP protein.

      In one embodiment, variants of a CRSP protein which function as either CRSP agonists (mimetics) or as CRSP antagonists can be identified by screening combinatorial  
20   libraries of mutants, e.g., truncation mutants, of a CRSP protein for CRSP protein agonist or antagonist activity. In one embodiment, a variegated library of CRSP variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of CRSP variants can be produced by, for  
25   example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential CRSP sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (e.g., for phage display) containing the set of CRSP sequences therein. There are a variety of methods which can be used to produce libraries of potential CRSP variants from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be  
30   performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an appropriate expression vector. Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding the desired set of potential CRSP sequences. Methods for synthesizing degenerate oligonucleotides are known in the art (see, e.g., Narang, S.A. (1983) *Tetrahedron* 39:3; Itakura et al. (1984) *Annu. Rev.*  
35   *Biochem.* 53:323; Itakura et al. (1984) *Science* 198:1056; Ike et al. (1983) *Nucleic Acid Res.* 11:477.

      In addition, libraries of fragments of a CRSP protein coding sequence can be used to generate a variegated population of CRSP fragments for screening and

subsequent selection of variants of a CRSP protein. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of a CRSP coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double stranded DNA which can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, an expression library can be derived which encodes N-terminal, C-terminal and internal fragments of various sizes of the CRSP protein.

Several techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Such techniques are adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of CRSP proteins. The most widely used techniques, which are amenable to high throughput analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a new technique which enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify CRSP variants (Arkin and Yourvan (1992) *PNAS* 89:7811-7815; Delgrave et al. (1993) *Protein Engineering* 6(3):327-331).

In one embodiment, cell based assays can be exploited to analyze a variegated CRSP library. For example, a library of expression vectors can be transfected into a cell line which ordinarily responds to a particular ligand in a CRSP-dependent manner. The transfected cells are then contacted with the ligand and the effect of expression of the mutant on signaling by the ligand can be detected, e.g., by measuring any of a number of immune cell responses. Plasmid DNA can then be recovered from the cells which score for inhibition, or alternatively, potentiation of ligand induction, and the individual clones further characterized.

An isolated CRSP protein, or a portion or fragment thereof, can be used as an immunogen to generate antibodies that bind CRSP using standard techniques for polyclonal and monoclonal antibody preparation. A full-length CRSP protein can be used or, alternatively, the invention provides antigenic peptide fragments of CRSP for use as immunogens. The antigenic peptide of CRSP comprises at least 8 amino acid residues of the amino acid sequence shown in SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8, or SEQ ID NO:11 and encompasses an epitope of CRSP such that an antibody raised against the peptide forms a specific immune complex with CRSP. Preferably, the

antigenic peptide comprises at least 10 amino acid residues, more preferably at least 15 amino acid residues, even more preferably at least 20 amino acid residues, and most preferably at least 30 amino acid residues.

Preferred epitopes encompassed by the antigenic peptide are regions of CRSP that are located on the surface of the protein, e.g., hydrophilic regions.

A CRSP immunogen typically is used to prepare antibodies by immunizing a suitable subject, (e.g., rabbit, goat, mouse or other mammal) with the immunogen. An appropriate immunogenic preparation can contain, for example, recombinantly expressed CRSP protein or a chemically synthesized CRSP polypeptide. The preparation can further include an adjuvant, such as Freund's complete or incomplete adjuvant, or similar immunostimulatory agent. Immunization of a suitable subject with an immunogenic CRSP preparation induces a polyclonal anti-CRSP antibody response.

Accordingly, another aspect of the invention pertains to anti-CRSP antibodies. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin molecules, i.e., molecules that contain an antigen binding site which specifically binds (immunoreacts with) an antigen, such as CRSP. Examples of immunologically active portions of immunoglobulin molecules include F(ab) and F(ab')<sub>2</sub> fragments which can be generated by treating the antibody with an enzyme such as pepsin. The invention provides polyclonal and monoclonal antibodies that bind CRSP. The term "monoclonal antibody" or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one species of an antigen binding site capable of immunoreacting with a particular epitope of CRSP. A monoclonal antibody composition thus typically displays a single binding affinity for a particular CRSP protein with which it immunoreacts.

Polyclonal anti-CRSP antibodies can be prepared as described above by immunizing a suitable subject with a CRSP immunogen. The anti-CRSP antibody titer in the immunized subject can be monitored over time by standard techniques, such as with an enzyme linked immunosorbent assay (ELISA) using immobilized CRSP. If desired, the antibody molecules directed against CRSP can be isolated from the mammal (e.g., from the blood) and further purified by well known techniques, such as protein A chromatography to obtain the IgG fraction. At an appropriate time after immunization, e.g., when the anti-CRSP antibody titers are highest, antibody-producing cells can be obtained from the subject and used to prepare monoclonal antibodies by standard techniques, such as the hybridoma technique originally described by Kohler and Milstein (1975) *Nature* 256:495-497) (see also, Brown et al. (1981) *J. Immunol.* 127:539-46; Brown et al. (1980) *J. Biol. Chem.* 255:4980-83; Yeh et al. (1976) *PNAS* 76:2927-31; and Yeh et al. (1982) *Int. J. Cancer* 29:269-75), the more recent human B



cell hybridoma technique (Kozbor et al. (1983) *Immunol Today* 4:72), the EBV-hybridoma technique (Cole et al. (1985), *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96) or trioma techniques. The technology for producing monoclonal antibody hybridomas is well known (see generally R. H. Kenneth, in  
5 *Monoclonal Antibodies: A New Dimension In Biological Analyses*, Plenum Publishing Corp., New York, New York (1980); E. A. Lerner (1981) *Yale J. Biol. Med.*, 54:387-402; M. L. Gefter et al. (1977) *Somatic Cell Genet.* 3:231-36). Briefly, an immortal cell line (typically a myeloma) is fused to lymphocytes (typically splenocytes) from a mammal immunized with a CRSP immunogen as described above, and the  
10 culture supernatants of the resulting hybridoma cells are screened to identify a hybridoma producing a monoclonal antibody that binds CRSP.

Any of the many well known protocols used for fusing lymphocytes and immortalized cell lines can be applied for the purpose of generating an anti-CRSP monoclonal antibody (see, e.g., G. Galfre et al. (1977) *Nature* 266:55052; Gefter et al.  
15 *Somatic Cell Genet.*, cited *supra*; Lerner, *Yale J. Biol. Med.*, cited *supra*; Kenneth, *Monoclonal Antibodies*, cited *supra*). Moreover, the ordinarily skilled worker will appreciate that there are many variations of such methods which also would be useful. Typically, the immortal cell line (e.g., a myeloma cell line) is derived from the same mammalian species as the lymphocytes. For example, murine hybridomas can be made  
20 by fusing lymphocytes from a mouse immunized with an immunogenic preparation of the present invention with an immortalized mouse cell line. Preferred immortal cell lines are mouse myeloma cell lines that are sensitive to culture medium containing hypoxanthine, aminopterin and thymidine ("HAT medium"). Any of a number of myeloma cell lines can be used as a fusion partner according to standard techniques,  
25 e.g., the P3-NS1/1-Ag4-1, P3-x63-Ag8.653 or Sp2/O-Ag14 myeloma lines. These myeloma lines are available from ATCC. Typically, HAT-sensitive mouse myeloma cells are fused to mouse splenocytes using polyethylene glycol ("PEG"). Hybridoma cells resulting from the fusion are then selected using HAT medium, which kills unfused and unproductively fused myeloma cells (unfused splenocytes die after several days  
30 because they are not transformed). Hybridoma cells producing a monoclonal antibody of the invention are detected by screening the hybridoma culture supernatants for antibodies that bind CRSP, e.g., using a standard ELISA assay.

Alternative to preparing monoclonal antibody-secreting hybridomas, a monoclonal anti-CRSP antibody can be identified and isolated by screening a  
35 recombinant combinatorial immunoglobulin library (e.g., an antibody phage display library) with CRSP to thereby isolate immunoglobulin library members that bind CRSP. Kits for generating and screening phage display libraries are commercially available (e.g., the Pharmacia *Recombinant Phage Antibody System*, Catalog No. 27-9400-01; and

- the Stratagene *SurfZAP™ Phage Display Kit*, Catalog No. 240612). Additionally, examples of methods and reagents particularly amenable for use in generating and screening antibody display library can be found in, for example, Ladner et al. U.S. Patent No. 5,223,409; Kang et al. PCT International Publication No. WO 92/18619; 5 Dower et al. PCT International Publication No. WO 91/17271; Winter et al. PCT International Publication WO 92/20791; Markland et al. PCT International Publication No. WO 92/15679; Breitling et al. PCT International Publication WO 93/01288; McCafferty et al. PCT International Publication No. WO 92/01047; Garrard et al. PCT International Publication No. WO 92/09690; Ladner et al. PCT International Publication 10 No. WO 90/02809; Fuchs et al. (1991) *Bio/Technology* 9:1370-1372; Hay et al. (1992) *Hum. Antibod. Hybridomas* 3:81-85; Huse et al. (1989) *Science* 246:1275-1281; Griffiths et al. (1993) *EMBO J* 12:725-734; Hawkins et al. (1992) *J. Mol. Biol.* 226:889-896; Clarkson et al. (1991) *Nature* 352:624-628; Gram et al. (1992) *PNAS* 89:3576-3580; Garrad et al. (1991) *Bio/Technology* 9:1373-1377; Hoogenboom et al. (1991) *Nuc.* 15 *Acid Res.* 19:4133-4137; Barbas et al. (1991) *PNAS* 88:7978-7982; and McCafferty et al. *Nature* (1990) 348:552-554.

- Additionally, recombinant anti-CRSP antibodies, such as chimeric and humanized monoclonal antibodies, comprising both human and non-human portions, which can be made using standard recombinant DNA techniques, are within the scope of 20 the invention. Such chimeric and humanized monoclonal antibodies can be produced by recombinant DNA techniques known in the art, for example using methods described in Robinson et al. International Application No. PCT/US86/02269; Akira, et al. European Patent Application 184,187; Taniguchi, M., European Patent Application 171,496; Morrison et al. European Patent Application 173,494; Neuberger et al. PCT International 25 Publication No. WO 86/01533; Cabilly et al. U.S. Patent No. 4,816,567; Cabilly et al. European Patent Application 125,023; Better et al. (1988) *Science* 240:1041-1043; Liu et al. (1987) *PNAS* 84:3439-3443; Liu et al. (1987) *J. Immunol.* 139:3521-3526; Sun et al. (1987) *PNAS* 84:214-218; Nishimura et al. (1987) *Canc. Res.* 47:999-1005; Wood et al. (1985) *Nature* 314:446-449; and Shaw et al. (1988) *J. Natl. Cancer Inst.* 80:1553- 30 1559; Morrison, S. L. (1985) *Science* 229:1202-1207; Oi et al. (1986) *BioTechniques* 4:214; Winter U.S. Patent 5,225,539; Jones et al. (1986) *Nature* 321:552-525; Verhoeven et al. (1988) *Science* 239:1534; and Beidler et al. (1988) *J. Immunol.* 141:4053-4060.

- An anti-CRSP antibody (e.g., monoclonal antibody) can be used to isolate CRSP 35 by standard techniques, such as affinity chromatography or immunoprecipitation. An anti-CRSP antibody can facilitate the purification of natural CRSP from cells and of recombinantly produced CRSP expressed in host cells. Moreover, an anti-CRSP antibody can be used to detect CRSP protein (e.g., in a cellular lysate or cell

supernatant) in order to evaluate the abundance and pattern of expression of the CRSP protein. Anti-CRSP antibodies can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, e.g., to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (i.e., physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase,  $\beta$ -galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include  $^{125}\text{I}$ ,  $^{131}\text{I}$ ,  $^{35}\text{S}$  or  $^3\text{H}$ .

### III. Recombinant Expression Vectors and Host Cells

Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding CRSP (or a portion thereof). As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g., bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (e.g., replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which

means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, which is operatively linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner which allows for expression of the nucleotide sequence (e.g., in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell). The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements (e.g., polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel; *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Regulatory sequences include those which direct constitutive expression of a nucleotide sequence in many types of host cell and those which direct expression of the nucleotide sequence only in certain host cells (e.g., tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (e.g., CRSP proteins, mutant forms of CRSP, fusion proteins, etc.).

The recombinant expression vectors of the invention can be designed for expression of CRSP in prokaryotic or eukaryotic cells. For example, CRSP can be expressed in bacterial cells such as *E. coli*, insect cells (using baculovirus expression vectors) yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out in *E. coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant protein; 2) to increase the solubility of the recombinant protein; and 3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase.

Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith, D.B. and Johnson, K.S. (1988) *Gene* 67:31-40), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

5 Purified fusion proteins can be utilized in CRSP activity assays, in CRSP ligand binding (e.g., direct assays or competitive assays described in detail below), to generate antibodies specific for CRSP proteins, as examples. In a preferred embodiment, a CRSP fusion expressed in a retroviral expression vector of the present invention can be utilized to infect bone marrow cells which are subsequently transplanted into irradiated  
10 recipients. The pathology of the subject recipient is then examined after sufficient time has passed (e.g six (6) weeks).

Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann et al., (1988) *Gene* 69:301-315) and pET 11d (Studier et al., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 60-89). Target gene expression from the pTrc vector relies on host  
15 RNA polymerase transcription from a hybrid trp-lac fusion promoter. Target gene expression from the pET 11d vector relies on transcription from a T7 gn10-lac fusion promoter mediated by a coexpressed viral RNA polymerase (T7 gn1). This viral polymerase is supplied by host strains BL21(DE3) or HMS174(DE3) from a resident  $\lambda$   
20 prophage harboring a T7 gn1 gene under the transcriptional control of the lacUV 5 promoter.

One strategy to maximize recombinant protein expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein (Gottesman, S., *Gene Expression Technology: Methods in*  
25 *Enzymology* 185, Academic Press, San Diego, California (1990) 119-128). Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (Wada et al., (1992) *Nucleic Acids Res.* 20:2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard  
30 DNA synthesis techniques.

In another embodiment, the CRSP expression vector is a yeast expression vector. Examples of vectors for expression in yeast *S. cerevisiae* include pYepSec1 (Baldari, et al., (1987) *Embo J.* 6:229-234), pMFa (Kurjan and Herskowitz, (1982) *Cell* 30:933-943), pJRY88 (Schultz et al., (1987) *Gene* 54:113-123), pYES2 (Invitrogen Corporation, San Diego, CA), and picZ (Invitrogen Corp, San Diego, CA).  
35

Alternatively, CRSP can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured

insect cells (e.g., Sf 9 cells) include the pAc series (Smith et al. (1983) *Mol. Cell Biol.* 3:2156-2165) and the pVL series (Lucklow and Summers (1989) *Virology* 170:31-39).

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, B. (1987) *Nature* 329:840) and pMT2PC (Kaufman et al. (1987) *EMBO J.* 6:187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells see chapters 16 and 17 of Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.*

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (e.g., tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert et al. (1987) *Genes Dev.* 1:268-277), lymphoid-specific promoters (Calame and Eaton (1988) *Adv. Immunol.* 43:235-275), in particular promoters of T cell receptors (Winoto and Baltimore (1989) *EMBO J.* 8:729-733) and immunoglobulins (Banerji et al. (1983) *Cell* 33:729-740; Queen and Baltimore (1983) *Cell* 33:741-748), neuron-specific promoters (e.g., the neurofilament promoter; Byrne and Ruddle (1989) *PNAS* 86:5473-5477), pancreas-specific promoters (Edlund et al. (1985) *Science* 230:912-916), and mammary gland-specific promoters (e.g., milk whey promoter; U.S. Patent No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, for example the murine hox promoters (Kessel and Gruss (1990) *Science* 249:374-379) and the  $\alpha$ -fetoprotein promoter (Campes and Tilghman (1989) *Genes Dev.* 3:537-546).

The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively linked to a regulatory sequence in a manner which allows for expression (by transcription of the DNA molecule) of an RNA molecule which is antisense to CRSP mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense orientation can be chosen which direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen which direct constitutive, tissue specific or cell type specific expression of antisense RNA. The

antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes see Weintraub, H. et al., Antisense RNA as a molecular tool for genetic analysis, *Reviews - Trends in Genetics*, Vol. 1(1) 1986.

Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, CRSP protein can be expressed in bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (e.g., DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, et al. (*Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989*), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (e.g., resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Preferred selectable markers include those which confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding CRSP or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (e.g., cells that have incorporated the selectable marker gene will survive, while the other cells die).

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (i.e., express) CRSP protein. Accordingly, the invention further provides methods for producing CRSP protein using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding CRSP has been introduced) in a suitable medium such that CRSP protein is produced. In another embodiment, the method further comprises isolating CRSP from the medium or the host cell.

The host cells of the invention can also be used to produce nonhuman transgenic animals. For example, in one embodiment, a host cell of the invention is a fertilized oocyte or an embryonic stem cell into which CRSP-coding sequences have been introduced. Such host cells can then be used to create non-human transgenic animals in which exogenous CRSP sequences have been introduced into their genome or homologous recombinant animals in which endogenous CRSP sequences have been altered. Such animals are useful for studying the function and/or activity of CRSP and for identifying and/or evaluating modulators of CRSP activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, etc. A transgene is exogenous DNA which is integrated into the genome of a cell from which a transgenic animal develops and which remains in the genome of the mature animal, thereby directing the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal. As used herein, a "homologous recombinant animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous CRSP gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, e.g., an embryonic cell of the animal, prior to development of the animal.

A transgenic animal of the invention can be created by introducing CRSP-encoding nucleic acid into the male pronuclei of a fertilized oocyte, e.g., by microinjection, retroviral infection, and allowing the oocyte to develop in a pseudopregnant female foster animal. The human CRSP cDNA sequence of SEQ ID NO:1, SEQ ID NO:4, SEQ ID NO:7, or SEQ ID NO:10 can be introduced as a transgene into the genome of a non-human animal. Alternatively, a nonhuman homologue of a human CRSP gene, such as a mouse CRSP gene, can be isolated based on hybridization to the human CRSP cDNA (described further in subsection I above) and used as a transgene. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. A tissue-specific



regulatory sequence(s) can be operably linked to the CRSP transgene to direct expression of CRSP protein to particular cells. Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009, both by Leder et al., U.S. Patent No. 4,873,191 by Wagner et al. and in Hogan, B., *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986). Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the CRSP transgene in its genome and/or expression of CRSP mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene encoding CRSP can further be bred to other transgenic animals carrying other transgenes.

To create a homologous recombinant animal, a vector is prepared which contains at least a portion of a CRSP gene into which a deletion, addition or substitution has been introduced to thereby alter, e.g., functionally disrupt, the CRSP gene. The CRSP gene can be a human gene (e.g., the cDNA of SEQ ID NO:3, SEQ ID NO: 6, SEQ ID NO:9 or SEQ ID NO:12), but more preferably, is a non-human homologue of a human CRSP gene. For example, a mouse CRSP gene of SEQ ID NO:16 can be used to construct a homologous recombination vector suitable for altering an endogenous CRSP gene in the mouse genome. In a preferred embodiment, the vector is designed such that, upon homologous recombination, the endogenous CRSP gene is functionally disrupted (i.e., no longer encodes a functional protein; also referred to as a "knock out" vector). Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous CRSP gene is mutated or otherwise altered but still encodes functional protein (e.g., the upstream regulatory region can be altered to thereby alter the expression of the endogenous CRSP protein). In the homologous recombination vector, the altered portion of the CRSP gene is flanked at its 5' and 3' ends by additional nucleic acid of the CRSP gene to allow for homologous recombination to occur between the exogenous CRSP gene carried by the vector and an endogenous CRSP gene in an embryonic stem cell. The additional flanking CRSP nucleic acid is of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5' and 3' ends) are included in the vector (see e.g., Thomas, K.R. and Capecchi, M. R. (1987) *Cell* 51:503 for a description of homologous recombination vectors). The vector is introduced into an embryonic stem cell line (e.g., by electroporation) and cells in which the introduced CRSP gene has homologously recombined with the endogenous CRSP gene are selected (see e.g., Li, E. et al. (1992) *Cell* 69:915). The selected cells are then injected into a blastocyst of an

animal (e.g., a mouse) to form aggregation chimeras (see e.g., Bradley, A. in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, E.J. Robertson, ed. (IRL, Oxford, 1987) pp. 113-152). A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny  
5 harboring the homologously recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously recombined DNA by germline transmission of the transgene. Methods for constructing homologous recombination vectors and homologous recombinant animals are described further in Bradley, A. (1991) *Current Opinion in Biotechnology* 2:823-829 and in PCT  
10 International Publication Nos.: WO 90/11354 by Le Mouellec et al.; WO 91/01140 by Smithies et al.; WO 92/0968 by Zijlstra et al.; and WO 93/04169 by Berns et al.

In another embodiment, transgenic non-humans animals can be produced which contain selected systems which allow for regulated expression of the transgene. One example of such a system is the *cre/loxP* recombinase system of bacteriophage P1. For  
15 a description of the *cre/loxP* recombinase system, see, e.g., Lakso et al. (1992) *PNAS* 89:6232-6236. Another example of a recombinase system is the FLP recombinase system of *Saccharomyces cerevisiae* (O'Gorman et al. (1991) *Science* 251:1351-1355. If a *cre/loxP* recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the *Cre* recombinase and a selected protein are  
20 required. Such animals can be provided through the construction of "double" transgenic animals, e.g., by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut, I. et al. (1997) *Nature* 385:810-  
25 813. In brief, a cell, e.g., a somatic cell, from the transgenic animal can be isolated and induced to exit the growth cycle and enter G<sub>0</sub> phase. The quiescent cell can then be fused, e.g., through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyte and then transferred to  
30 pseudopregnant female foster animal. The offspring borne of this female foster animal will be a clone of the animal from which the cell, e.g., the somatic cell, is isolated.

#### IV. Pharmaceutical Compositions

The CRSP nucleic acid molecules, CRSP proteins, and anti-CRSP antibodies  
35 (also referred to herein as "active compounds") of the invention can be incorporated into pharmaceutical compositions suitable for administration. Such compositions typically comprise the nucleic acid molecule, protein, or antibody and a pharmaceutically acceptable carrier. As used herein the language "pharmaceutically acceptable carrier" is

intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, e.g., intravenous, intradermal, subcutaneous, oral (e.g., inhalation), transdermal (topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, NJ) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as manitol, sorbitol, sodium

chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound (e.g., a CRSP protein or anti-CRSP antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle which contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and freeze-drying which yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, e.g., a gas such as carbon dioxide, or a nebulizer.

Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

The compounds can also be prepared in the form of suppositories (e.g., with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

5 In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art.

10 The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

15 It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required

20 pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

Toxicity and therapeutic efficacy of such compounds can be determined by

25 standard pharmaceutical procedures in cell cultures or experimental animals, *e.g.*, for determining the LD50 (the dose lethal to 50% of the population) and the ED50 (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD50/ED50. Compounds which exhibit large therapeutic indices are preferred. While

30 compounds that exhibit toxic side effects may be used, care should be taken to design a delivery system that targets such compounds to the site of affected tissue in order to minimize potential damage to uninfected cells and, thereby, reduce side effects.

The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such compounds lies

35 preferably within a range of circulating concentrations that include the ED50 with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially

from cell culture assays. A dose may be formulated in animal models to achieve a circulating plasma concentration range that includes the IC<sub>50</sub> (i.e., the concentration of the test compound which achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma may be measured, for example, by high performance liquid chromatography.

The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (see U.S. Patent 5,328,470) or by stereotactic injection (see e.g., Chen et al. (1994) *PNAS* 91:3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, e.g., retroviral vectors, the pharmaceutical preparation can include one or more cells which produce the gene delivery system.

The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

## V. Uses and Methods of the Invention

The nucleic acid molecules, proteins, protein homologues, and antibodies described herein can be used in one or more of the following methods: a) screening assays; b) predictive medicine (e.g., diagnostic assays, prognostic assays, monitoring clinical trials, and pharmacogenetics); and c) methods of treatment (e.g., therapeutic and prophylactic). As described herein, a CRSP protein of the invention has one or more of the following activities: (i) interaction of a CRSP protein with a membrane-bound CRSP receptor; (ii) interaction of a CRSP protein with a soluble form of a CRSP receptor; (iii) interaction of a CRSP protein with an intracellular protein via a membrane-bound CRSP receptor; (iv) complex formation between a soluble CRSP protein and a second soluble CRSP binding partner; (iv) complex formation between a soluble CRSP protein and a second soluble CRSP binding partner, wherein the CRSP binding partner is a non-CRSP protein molecule; and (v) complex formation between a soluble CRSP protein and a second soluble CRSP binding partner, wherein the CRSP binding partner is a second CRSP protein molecule and can thus be used in, for example, (1) modulation of cellular signal transduction, either *in vitro* or *in vivo*; (2) regulation of gene transcription in a cell involved in development or differentiation, either *in vitro* or *in vivo*; (3) regulation of gene transcription in a cell involved in development or differentiation, wherein at least one gene encodes a differentiation-

specific protein; (4) regulation of gene transcription in a cell involved in development or differentiation, wherein at least one gene encodes a second secreted protein; (5) regulation of gene transcription in a cell involved in development or differentiation, wherein at least one gene encodes a signal transduction molecule; and (6) regulation of cellular proliferation, either *in vitro* or *in vivo*. The isolated nucleic acid molecules of the invention can be used, for example, to express CRSP protein (e.g., via a recombinant expression vector in a host cell in gene therapy applications), to detect CRSP mRNA (e.g., in a biological sample) or a genetic alteration in a CRSP gene, and to modulate CRSP activity, as described further below. In addition, the CRSP proteins can be used to screen drugs or compounds which modulate the CRSP activity as well as to treat disorders characterized by insufficient or excessive production of CRSP protein or production of CRSP protein forms which have decreased or aberrant activity compared to CRSP wild type protein (e.g., developmental disorders or proliferative diseases such as cancer). Moreover, the anti-CRSP antibodies of the invention can be used to detect and isolate CRSP proteins, regulate the bioavailability of CRSP proteins, and modulate CRSP activity.

#### A. Screening Assays:

The invention provides a method (also referred to herein as a "screening assay") for identifying modulators, i.e., candidate or test compounds or agents (e.g., peptides, peptidomimetics, small molecules or other drugs) which bind to CRSP proteins or have a stimulatory or inhibitory effect on, for example, CRSP expression or CRSP activity.

In one embodiment, the invention provides assays for screening candidate or test compounds which bind to or modulate the activity of a CRSP protein or polypeptide or biologically active portion thereof. In another embodiment, the invention provides assays for screening candidate or test compounds which bind to or modulate the activity of a CRSP receptor. The test compounds of the present invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the 'one-bead one-compound' library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds (Lam, K.S. (1997) *Anticancer Drug Des.* 12:145).

Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt *et al.* (1993) *Proc. Natl. Acad. Sci. U.S.A.* 90:6909; Erb *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:11422; Zuckermann *et al.* (1994). *J. Med.*

*Chem.* 37:2678; Cho et al. (1993) *Science* 261:1303; Carrell et al. (1994) *Angew. Chem. Int. Ed. Engl.* 33:2059; Carrell et al. (1994) *Angew. Chem. Int. Ed. Engl.* 33:2061; and in Gallop et al. (1994) *J. Med. Chem.* 37:1233.

Libraries of compounds may be presented in solution (e.g., Houghten (1992) *Biotechniques* 13:412-421), or on beads (Lam (1991) *Nature* 354:82-84), chips (Fodor (1993) *Nature* 364:555-556), bacteria (Ladner USP 5,223,409), spores (Ladner USP '409), plasmids (Cull et al. (1992) *Proc Natl Acad Sci USA* 89:1865-1869) or on phage (Scott and Smith (1990) *Science* 249:386-390); (Devlin (1990) *Science* 249:404-406); (Cwirla et al. (1990) *Proc. Natl. Acad. Sci.* 87:6378-6382); (Felici (1991) *J. Mol. Biol.* 222:301-310); (Ladner *supra.*).

In one embodiment, an assay is a cell-based assay in which a cell which expresses a CRSP receptor on the cell surface is contacted with a test compound and the ability of the test compound to bind to a CRSP receptor determined. The cell, for example, can be of mammalian origin or a yeast cell. Determining the ability of the test compound to bind to a CRSP receptor can be accomplished, for example, by coupling the test compound with a radioisotope or enzymatic label such that binding of the test compound to the CRSP receptor can be determined by detecting the labeled compound in a complex. For example, test compounds can be labeled with  $^{125}\text{I}$ ,  $^{35}\text{S}$ ,  $^{14}\text{C}$ , or  $^3\text{H}$ , either directly or indirectly, and the radioisotope detected by direct counting of radioemmission or by scintillation counting. Alternatively, test compounds can be enzymatically labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product.

It is also within the scope of this invention to determine the ability of a test compound to interact with a CRSP receptor without the labeling of any of the interactants. For example, a microphysiometer can be used to detect the interaction of a test compound with a CRSP receptor without the labeling of either the test compound or the receptor. McConnell, H. M. et al. (1992) *Science* 257:1906-1912. As used herein, a "microphysiometer" (e.g., Cytosensor™) is an analytical instrument that measures the rate at which a cell acidifies its environment using a light-addressable potentiometric sensor (LAPS). Changes in this acidification rate can be used as an indicator of the interaction between ligand and receptor.

In a preferred embodiment, the assay comprises contacting a cell which expresses a CRSP receptor on the cell surface with a CRSP protein or biologically-active portion thereof, to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with an CRSP receptor, wherein determining the ability of the test compound to interact with an CRSP receptor comprises determining the ability of the test compound to preferentially bind to



the CRSP receptor as compared to the ability of CRSP, or a biologically active portion thereof, to bind to the receptor.

In another embodiment, an assay is a cell-based assay comprising contacting a cell expressing a CRSP target molecule with a test compound and determining the ability of the test compound to modulate (e.g. stimulate or inhibit) the activity of the CRSP target molecule. Determining the ability of the test compound to modulate the activity of a CRSP target molecule can be accomplished, for example, by determining the ability of the CRSP protein to bind to or interact with the CRSP target molecule.

Determining the ability of the CRSP protein to bind to or interact with a CRSP target molecule can be accomplished by one of the methods described above for determining direct binding. In a preferred embodiment, determining the ability of the CRSP protein to bind to or interact with an CRSP target molecule can be accomplished by determining the activity of the target molecule. For example, the activity of the target molecule can be determined by detecting induction of a cellular second messenger of the target (i.e. intracellular  $\text{Ca}^{2+}$ , diacylglycerol,  $\text{IP}_3$ , etc.), detecting catalytic/enzymatic activity of the target an appropriate substrate, detecting the induction of a reporter gene (comprising a CRSP-responsive regulatory element operatively linked to a nucleic acid encoding a detectable marker, e.g., luciferase), or detecting a cellular response, for example, development, differentiation or rate of proliferation.

In yet another embodiment, an assay of the present invention is a cell-free assay in which a CRSP protein or biologically active portion thereof is contacted with a test compound and the ability of the test compound to bind to the CRSP protein or biologically active portion thereof is determined. Binding of the test compound to the CRSP protein can be determined either directly or indirectly as described above. In a preferred embodiment, the assay includes contacting the CRSP protein or biologically active portion thereof with a known compound which binds CRSP to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a CRSP protein, wherein determining the ability of the test compound to interact with a CRSP protein comprises determining the ability of the test compound to preferentially bind to CRSP or biologically active portion thereof as compared to the known compound.

In another embodiment, the assay is a cell-free assay in which a CRSP protein or biologically active portion thereof is contacted with a test compound and the ability of the test compound to modulate (e.g., stimulate or inhibit) the activity of the CRSP protein or biologically active portion thereof is determined. Determining the ability of the test compound to modulate the activity of a CRSP protein can be accomplished, for example, by determining the ability of the CRSP protein to bind to a CRSP target molecule by one of the methods described above for determining direct binding.

Determining the ability of the CRSP protein to bind to a CRSP target molecule can also be accomplished using a technology such as real-time Biomolecular Interaction Analysis (BIA). Sjolander, S. and Urbaniczky, C. (1991) *Anal. Chem.* 63:2338-2345 and Szabo et al. (1995) *Curr. Opin. Struct. Biol.* 5:699-705. As used herein, "BIA" is a technology  
5 for studying biospecific interactions in real time, without labeling any of the interactants (e.g., BIAcore™). Changes in the optical phenomenon surface plasmon resonance (SPR) can be used as an indication of real-time reactions between biological molecules.

In an alternative embodiment, determining the ability of the test compound to modulate the activity of a CRSP protein can be accomplished by determining the ability  
10 of the CRSP protein to further modulate the activity of a CRSP target molecule. For example, the catalytic/enzymatic activity of the target molecule on an appropriate substrate can be determined as previously described.

In yet another embodiment, the cell-free assay involves contacting a CRSP protein or biologically active portion thereof with a known compound which binds the  
15 CRSP protein to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with the CRSP protein, wherein determining the ability of the test compound to interact with the CRSP protein comprises determining the ability of the CRSP protein to preferentially bind to or modulate the activity of a CRSP target molecule.

20 The cell-free assays of the present invention are amenable to use of both soluble and/or membrane-bound forms of isolated proteins (e.g. CRSP proteins or biologically active portions thereof or CRSP target molecules). In the case of cell-free assays in which a membrane-bound form an isolated protein is used (e.g., an CRSP target molecule or receptor) it may be desirable to utilize a solubilizing agent such that the  
25 membrane-bound form of the isolated protein is maintained in solution. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-dodecylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton® X-100, Triton® X-114, Thesit®, Isotridecypoly(ethylene glycol ether)<sub>n</sub>, 3-[(3-cholamidopropyl)dimethylamminio]-1-propane sulfonate (CHAPS),  
30 3-[(3-cholamidopropyl)dimethylamminio]-2-hydroxy-1-propane sulfonate (CHAPSO), or N-dodecyl=N,N-dimethyl-3-ammonio-1-propane sulfonate.

In more than one embodiment of the above assay methods of the present invention, it may be desirable to immobilize either CRSP or its target molecule to facilitate separation of complexed from uncomplexed forms of one or both of the  
35 proteins, as well as to accommodate automation of the assay. Binding of a test compound to a CRSP protein, or interaction of a CRSP protein with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtitre plates,

test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided which adds a domain that allows one or both of the proteins to be bound to a matrix. For example, glutathione-S-transferase/ CRSP fusion proteins or glutathione-S-transferase/target fusion proteins can be adsorbed onto glutathione sepharose beads  
5 (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtitre plates, which are then combined with the test compound or the test compound and either the non-adsorbed target protein or CRSP protein, and the mixture incubated under conditions conducive to complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads or microtitre plate wells are washed to remove any unbound  
10 components, the matrix immobilized in the case of beads, complex determined either directly or indirectly, for example, as described above. Alternatively, the complexes can be dissociated from the matrix, and the level of CRSP binding or activity determined using standard techniques.

Other techniques for immobilizing proteins on matrices can also be used in the  
15 screening assays of the invention. For example, either a CRSP protein or a CRSP target molecule can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated CRSP protein or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques well known in the art (e.g., biotinylation kit, Pierce Chemicals, Rockford, IL), and immobilized in the wells of streptavidin-coated 96  
20 well plates (Pierce Chemical). Alternatively, antibodies reactive with CRSP protein or target molecules but which do not interfere with binding of the CRSP protein to its target molecule can be derivatized to the wells of the plate, and unbound target or CRSP protein trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes,  
25 include immunodetection of complexes using antibodies reactive with the CRSP protein or target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the CRSP protein or target molecule.

In another embodiment, modulators of CRSP expression are identified in a method wherein a cell is contacted with a candidate compound and the expression of  
30 CRSP mRNA or protein in the cell is determined. The level of expression of CRSP mRNA or protein in the presence of the candidate compound is compared to the level of expression of CRSP mRNA or protein in the absence of the candidate compound. The candidate compound can then be identified as a modulator of CRSP expression based on this comparison. For example, when expression of CRSP mRNA or protein is greater  
35 (statistically significantly greater) in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of CRSP mRNA or protein expression. Alternatively, when expression of CRSP mRNA or protein is less (statistically significantly less) in the presence of the candidate compound than in its

absence, the candidate compound is identified as an inhibitor of CRSP mRNA or protein expression. The level of CRSP mRNA or protein expression in the cells can be determined by methods described herein for detecting CRSP mRNA or protein.

In yet another aspect of the invention, the CRSP proteins can be used as "bait proteins" in a two-hybrid assay or three-hybrid assay (see, e.g., U.S. Patent No. 5,283,317; Zervos *et al.* (1993) *Cell* 72:223-232; Madura *et al.* (1993) *J. Biol. Chem.* 268:12046-12054; Bartel *et al.* (1993) *Biotechniques* 14:920-924; Iwabuchi *et al.* (1993) *Oncogene* 8:1693-1696; and Brent WO94/10300), to identify other proteins, which bind to or interact with CRSP ("CRSP-binding proteins" or "CRSP-bp") and modulate CRSP activity. Such CRSP-binding proteins are also likely to be involved in the propagation of signals by the CRSP proteins as, for example, downstream elements of a CRSP-mediated signaling pathway. Alternatively, such CRSP-binding proteins are likely to be cell-surface molecules associated with non-CRSP expressing cells, wherein such CRSP-binding proteins are involved in signal transduction.

The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for a CRSP protein is fused to a gene encoding the DNA binding domain of a known transcription factor (e.g., GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, *in vivo*, forming an CRSP-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (e.g., LacZ) which is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene which encodes the protein which interacts with the CRSP protein.

This invention further pertains to novel agents identified by the above-described screening assays. Accordingly, it is within the scope of this invention to further use an agent identified as described herein in an appropriate animal model. For example, an agent identified as described herein (e.g., a CRSP modulating agent, an antisense CRSP nucleic acid molecule, a CRSP-specific antibody, or a CRSP-binding partner) can be used in an animal model to determine the efficacy, toxicity, or side effects of treatment with such an agent. Alternatively, an agent identified as described herein can be used in an animal model to determine the mechanism of action of such an agent. Furthermore, this invention pertains to uses of novel agents identified by the above-described screening assays for treatments as described herein.

## B. Detection Assays

Portions or fragments of the cDNA sequences identified herein (and the corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. For example, these sequences can be used to: (i) map their respective genes on a chromosome; and, thus, locate gene regions associated with genetic disease; (ii) identify an individual from a minute biological sample (tissue typing); and (iii) aid in forensic identification of a biological sample. These applications are described in the subsections below.

### 1. Chromosome Mapping

Once the sequence (or a portion of the sequence) of a gene has been isolated, this sequence can be used to map the location of the gene on a chromosome. This process is called chromosome mapping. Accordingly, portions or fragments of the CRSP nucleotide sequences, described herein, can be used to map the location of the CRSP genes on a chromosome. The mapping of the CRSP sequences to chromosomes is an important first step in correlating these sequences with genes associated with disease.

Briefly, CRSP genes can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp in length) from the CRSP nucleotide sequences. Computer analysis of the CRSP sequences can be used to predict primers that do not span more than one exon in the genomic DNA, thus complicating the amplification process. These primers can then be used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the CRSP sequences will yield an amplified fragment.

Somatic cell hybrids are prepared by fusing somatic cells from different mammals (e.g., human and mouse cells). As hybrids of human and mouse cells grow and divide, they gradually lose human chromosomes in random order, but retain the mouse chromosomes. By using media in which mouse cells cannot grow, because they lack a particular enzyme, but human cells can, the one human chromosome that contains the gene encoding the needed enzyme, will be retained. By using various media, panels of hybrid cell lines can be established. Each cell line in a panel contains either a single human chromosome or a small number of human chromosomes, and a full set of mouse chromosomes, allowing easy mapping of individual genes to specific human chromosomes. (D'Eustachio P. et al. (1983) *Science* 220:919-924). Somatic cell hybrids containing only fragments of human chromosomes can also be produced by using human chromosomes with translocations and deletions.

PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular sequence to a particular chromosome. Three or more sequences can be

assigned per day using a single thermal cycler. Using the CRSP nucleotide sequences to design oligonucleotide primers, sublocalization can be achieved with panels of fragments from specific chromosomes. Other mapping strategies which can similarly be used to map a 9o, 1p, or 1v sequence to its chromosome include *in situ* hybridization  
5 (described in Fan, Y. et al. (1990) *PNAS*, 87:6223-27), pre-screening with labeled flow-sorted chromosomes, and pre-selection by hybridization to chromosome specific cDNA libraries.

Fluorescence *in situ* hybridization (FISH) of a DNA sequence to a metaphase chromosomal spread can further be used to provide a precise chromosomal location in  
10 one step. Chromosome spreads can be made using cells whose division has been blocked in metaphase by a chemical such as colcemid that disrupts the mitotic spindle. The chromosomes can be treated briefly with trypsin, and then stained with Giemsa. A pattern of light and dark bands develops on each chromosome, so that the chromosomes can be identified individually. The FISH technique can be used with a DNA sequence  
15 as short as 500 or 600 bases. However, clones larger than 1,000 bases have a higher likelihood of binding to a unique chromosomal location with sufficient signal intensity for simple detection. Preferably 1,000 bases, and more preferably 2,000 bases will suffice to get good results at a reasonable amount of time. For a review of this technique, see Verma *et al.*, Human Chromosomes: A Manual of Basic Techniques  
20 (Pergamon Press, New York 1988).

Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding  
25 sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. (Such data are found, for example, in V. McKusick, Mendelian Inheritance in  
30 Man, available on-line through Johns Hopkins University Welch Medical Library). The relationship between a gene and a disease, mapped to the same chromosomal region, can then be identified through linkage analysis (co-inheritance of physically adjacent genes), described in, for example, Egeland, J. *et al.* (1987) *Nature*, 325:783-787.

Moreover, differences in the DNA sequences between individuals affected and  
35 unaffected with a disease associated with the CRSP gene, can be determined. If a mutation is observed in some or all of the affected individuals but not in any unaffected individuals, then the mutation is likely to be the causative agent of the particular disease. Comparison of affected and unaffected individuals generally involves first looking for

structural alterations in the chromosomes, such as deletions or translocations that are visible from chromosome spreads or detectable using PCR based on that DNA sequence. Ultimately, complete sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

5

## 2. Tissue Typing

The CRSP sequences of the present invention can also be used to identify individuals from minute biological samples. The United States military, for example, is considering the use of restriction fragment length polymorphism (RFLP) for  
10 identification of its personnel. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identification. This method does not suffer from the current limitations of "Dog Tags" which can be lost, switched, or stolen, making positive identification difficult. The sequences of the present invention are useful as additional DNA markers  
15 for RFLP (described in U.S. Patent 5,272,057).

Furthermore, the sequences of the present invention can be used to provide an alternative technique which determines the actual base-by-base DNA sequence of selected portions of an individual's genome. Thus, the CRSP nucleotide sequences described herein can be used to prepare two PCR primers from the 5' and 3' ends of the  
20 sequences. These primers can then be used to amplify an individual's DNA and subsequently sequence it.

Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences due to allelic differences. The sequences of the  
25 present invention can be used to obtain such identification sequences from individuals and from tissue. The CRSP nucleotide sequences of the invention uniquely represent portions of the human genome. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. It is estimated that allelic variation between individual humans occurs with a frequency of  
30 about once per each 500 bases. Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are necessary to differentiate individuals. The noncoding sequences of SEQ ID NO:1, SEQ ID NO:4, SEQ ID NO:7, or SEQ ID  
35 NO:10, can comfortably provide positive individual identification with a panel of perhaps 10 to 1,000 primers which each yield a noncoding amplified sequence of 100 bases. If predicted coding sequences, such as those in SEQ ID NO:3, SEQ ID NO:6,

SEQ ID NO:9, or SEQ ID NO:12 are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

If a panel of reagents from CRSP nucleotide sequences described herein is used to generate a unique identification database for an individual, those same reagents can  
5 later be used to identify tissue from that individual. Using the unique identification database, positive identification of the individual, living or dead, can be made from extremely small tissue samples.

### 3. Use of Partial CRSP Sequences in Forensic Biology

10 DNA-based identification techniques can also be used in forensic biology. Forensic biology is a scientific field employing genetic typing of biological evidence found at a crime scene as a means for positively identifying, for example, a perpetrator of a crime. To make such an identification, PCR technology can be used to amplify DNA sequences taken from very small biological samples such as tissues, e.g., hair or  
15 skin, or body fluids, e.g., blood, saliva, or semen found at a crime scene. The amplified sequence can then be compared to a standard, thereby allowing identification of the origin of the biological sample.

The sequences of the present invention can be used to provide polynucleotide reagents, e.g., PCR primers, targeted to specific loci in the human genome, which can  
20 enhance the reliability of DNA-based forensic identifications by, for example, providing another "identification marker" (i.e. another DNA sequence that is unique to a particular individual). As mentioned above, actual base sequence information can be used for identification as an accurate alternative to patterns formed by restriction enzyme generated fragments. Sequences targeted to noncoding regions of SEQ ID NOs:1, SEQ  
25 ID NO:4, SEQ ID NO:7, or SEQ ID NO:10 are particularly appropriate for this use as greater numbers of polymorphisms occur in the noncoding regions, making it easier to differentiate individuals using this technique. Examples of polynucleotide reagents include the CRSP nucleotide sequences or portions thereof, e.g., fragments derived from the noncoding regions of SEQ ID NO:1, SEQ ID NO:4, SEQ ID NO:7, or SEQ ID  
30 NO:10, having a length of at least 20 bases, preferably at least 30 bases.

The CRSP nucleotide sequences described herein can further be used to provide polynucleotide reagents, e.g., labeled or labelable probes which can be used in, for example, an *in situ* hybridization technique, to identify a specific tissue, e.g., brain  
35 tissue. This can be very useful in cases where a forensic pathologist is presented with a tissue of unknown origin. Panels of such CRSP probes can be used to identify tissue by species and/or by organ type.



In a similar fashion, these reagents, e.g., CRSP primers or probes can be used to screen tissue culture for contamination (i.e. screen for the presence of a mixture of different types of cells in a culture).

5 C. Predictive Medicine:

The present invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual prophylactically. Accordingly, one aspect of the present invention relates to diagnostic assays for determining CRSP protein and/or nucleic acid expression as well as CRSP activity, in the context of a biological sample (e.g., blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant CRSP expression or activity. The invention also provides for prognostic (or predictive) assays for determining whether an individual is at risk of developing a disorder associated with CRSP protein, nucleic acid expression or activity. For example, mutations in an CRSP gene can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of a disorder characterized by or associated with CRSP protein, nucleic acid expression or activity.

Another aspect of the invention pertains to monitoring the influence of agents (e.g., drugs, compounds) on the expression or activity of CRSP in clinical trials.

These and other agents are described in further detail in the following sections.

1. Diagnostic Assays

25 An exemplary method for detecting the presence or absence of CRSP protein or nucleic acid in a biological sample involves obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting CRSP protein or nucleic acid (e.g., mRNA, genomic DNA) that encodes CRSP protein such that the presence of CRSP protein or nucleic acid is detected in the biological sample. A preferred agent for detecting CRSP mRNA or genomic DNA is a labeled nucleic acid probe capable of hybridizing to CRSP mRNA or genomic DNA. The nucleic acid probe can be, for example, a full-length CRSP nucleic acid, such as the nucleic acid of SEQ ID NO: 1, SEQ ID NO:4, SEQ ID NO:7, SEQ ID NO:10, the DNA insert of the plasmid deposited with ATCC as Accession Number \_\_\_\_\_, the DNA insert of the plasmid deposited with ATCC as Accession Number 98633, or a portion thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to CRSP mRNA or

genomic DNA. Other suitable probes for use in the diagnostic assays of the invention are described herein.

A preferred agent for detecting CRSP protein is an antibody capable of binding to CRSP protein, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (e.g., Fab or F(ab')<sub>2</sub>) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (i.e., physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. That is, the detection method of the invention can be used to detect CRSP mRNA, protein, or genomic DNA in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection of CRSP mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of CRSP protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence. *In vitro* techniques for detection of CRSP genomic DNA include Southern hybridizations. Furthermore, *in vivo* techniques for detection of CRSP protein include introducing into a subject a labeled anti-CRSP antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

In one embodiment, the biological sample contains protein molecules from the test subject. Alternatively, the biological sample can contain mRNA molecules from the test subject or genomic DNA molecules from the test subject. A preferred biological sample is a serum sample isolated by conventional means from a subject.

In another embodiment, the methods further involve obtaining a control biological sample from a control subject, contacting the control sample with a compound or agent capable of detecting CRSP protein, mRNA, or genomic DNA, such that the presence of CRSP protein, mRNA or genomic DNA is detected in the biological sample, and comparing the presence of CRSP protein, mRNA or genomic DNA in the control sample with the presence of CRSP protein, mRNA or genomic DNA in the test sample.

The invention also encompasses kits for detecting the presence of CRSP in a biological sample. For example, the kit can comprise a labeled compound or agent capable of detecting CRSP protein or mRNA in a biological sample; means for

determining the amount of CRSP in the sample; and means for comparing the amount of CRSP in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect CRSP protein or nucleic acid.

5

## 2. Prognostic Assays

The diagnostic methods described herein can furthermore be utilized to identify subjects having or at risk of developing a disease or disorder associated with aberrant CRSP expression or activity. For example, the assays described herein, such as the preceding diagnostic assays or the following assays, can be utilized to identify a subject having or at risk of developing a disorder associated with CRSP protein, nucleic acid expression or activity such as a proliferative disorder, a differentiative or developmental disorder, or a hematopoietic disorder. Alternatively, the prognostic assays can be utilized to identify a subject having or at risk for developing a differentiative or proliferative disease (e.g., cancer). Thus, the present invention provides a method for identifying a disease or disorder associated with aberrant CRSP expression or activity in which a test sample is obtained from a subject and CRSP protein or nucleic acid (e.g., mRNA, genomic DNA) is detected, wherein the presence of CRSP protein or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant CRSP expression or activity. As used herein, a "test sample" refers to a biological sample obtained from a subject of interest. For example, a test sample can be a biological fluid (e.g., serum), cell sample, or tissue.

Furthermore, the prognostic assays described herein can be used to determine whether a subject can be administered an agent (e.g., an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant CRSP expression or activity. For example, such methods can be used to determine whether a subject can be effectively treated with an agent for a disorder, such as a proliferative disorder, a differentiative or developmental disorder, or a hematopoietic disorder. Alternatively, such methods can be used to determine whether a subject can be effectively treated with an agent for a differentiative or proliferative disease (e.g., cancer). Thus, the present invention provides methods for determining whether a subject can be effectively treated with an agent for a disorder associated with aberrant CRSP expression or activity in which a test sample is obtained and CRSP protein or nucleic acid expression or activity is detected (e.g., wherein the abundance of CRSP protein or nucleic acid expression or activity is diagnostic for a subject that can be administered the agent to treat a disorder associated with aberrant CRSP expression or activity.)

The methods of the invention can also be used to detect genetic alterations in an CRSP gene, thereby determining if a subject with the altered gene is at risk for a disorder characterized by aberrant development, aberrant cellular differentiation, aberrant cellular proliferation or an aberrant hematopoietic response. In preferred  
5 embodiments, the methods include detecting, in a sample of cells from the subject, the presence or absence of a genetic alteration characterized by at least one of an alteration affecting the integrity of a gene encoding a CRSP-protein, or the mis-expression of the CRSP gene. For example, such genetic alterations can be detected by ascertaining the existence of at least one of 1) a deletion of one or more nucleotides from an CRSP gene;  
10 2) an addition of one or more nucleotides to a CRSP gene; 3) a substitution of one or more nucleotides of a CRSP gene, 4) a chromosomal rearrangement of a CRSP gene; 5) an alteration in the level of a messenger RNA transcript of a CRSP gene, 6) aberrant modification of a CRSP gene, such as of the methylation pattern of the genomic DNA, 7) the presence of a non-wild type splicing pattern of a messenger RNA transcript of a  
15 CRSP gene, 8) a non-wild type level of a CRSP-protein, 9) allelic loss of a CRSP gene, and 10) inappropriate post-translational modification of a CRSP-protein. As described herein, there are a large number of assay techniques known in the art which can be used for detecting alterations in a CRSP gene. A preferred biological sample is a tissue or serum sample isolated by conventional means from a subject.

20 In certain embodiments, detection of the alteration involves the use of a probe/primer in a polymerase chain reaction (PCR) (see, e.g., U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (see, e.g., Landegran *et al.* (1988) *Science* 241:1077-1080; and Nakazawa *et al.* (1994) *PNAS* 91:360-364), the latter of which can be particularly  
25 useful for detecting point mutations in the CRSP-gene (see Abravaya *et al.* (1995) *Nucleic Acids Res.* 23:675-682). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (e.g., genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers which specifically hybridize to a CRSP gene under conditions such that  
30 hybridization and amplification of the CRSP-gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

35 Alternative amplification methods include: self sustained sequence replication (Guatelli, J.C. *et al.*, 1990, Proc. Natl. Acad. Sci. USA 87:1874-1878), transcriptional amplification system (Kwoh, D.Y. *et al.*, 1989, Proc. Natl. Acad. Sci. USA 86:1173-1177), Q-Beta Replicase (Lizardi, P.M. *et al.*, 1988, Bio/Technology 6:1197), or any

other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

5 In an alternative embodiment, mutations in a CRSP gene from a sample cell can be identified by alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (see, for example, U.S. Patent No. 5,498,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

In other embodiments, genetic mutations in CRSP can be identified by hybridizing a sample and control nucleic acids, e.g., DNA or RNA, to high density arrays containing hundreds or thousands of oligonucleotides probes (Cronin, M.T. *et al.* (1996) *Human Mutation* 7: 244-255; Kozal, M.J. *et al.* (1996) *Nature Medicine* 2: 753-759). For example, genetic mutations in CRSP can be identified in two dimensional arrays containing light-generated DNA probes as described in Cronin, M.T. *et al. supra*. Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This step is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the CRSP gene and detect mutations by comparing the sequence of the sample CRSP with the corresponding wild-type (control) sequence. Examples of sequencing reactions include those based on techniques developed by Maxim and Gilbert ((1977) *PNAS* 74:560) or Sanger ((1977) *PNAS* 74:5463). It is also contemplated that any of a variety of automated sequencing procedures can be utilized when performing the diagnostic assays ((1995) *Biotechniques* 19:448), including sequencing by mass spectrometry (see, e.g., PCT International Publication No. WO 94/16101; Cohen *et al.* (1996) *Adv. Chromatogr.* 36:127-162; and Griffin *et al.* (1993) *Appl. Biochem. Biotechnol.* 38:147-159).

Other methods for detecting mutations in the CRSP gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes (Myers *et al.* (1985) *Science* 230:1242). In general, the art technique of "mismatch cleavage" starts by providing heteroduplexes of formed by hybridizing (labeled) RNA or DNA containing the wild-type CRSP sequence with potentially mutant RNA or DNA obtained from a tissue sample. The double-stranded duplexes are treated with an agent which cleaves single-stranded regions of the duplex such as which will exist due to basepair mismatches between the control and sample strands. For instance, RNA/DNA duplexes can be treated with RNase and DNA/DNA hybrids treated with S1 nuclease to enzymatically digesting the mismatched regions. In other embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium tetroxide and with piperidine in order to digest mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine the site of mutation. See, for example, Cotton *et al.* (1988) *Proc. Natl Acad Sci USA* 85:4397; Saleeba *et al.* (1992) *Methods Enzymol.* 217:286-295. In a preferred embodiment, the control DNA or RNA can be labeled for detection.

In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair" enzymes) in defined systems for detecting and mapping point mutations in CRSP cDNAs obtained from samples of cells. For example, the mutY enzyme of *E. coli* cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches (Hsu *et al.* (1994) *Carcinogenesis* 15:1657-1662). According to an exemplary embodiment, a probe based on an CRSP sequence, e.g., a wild-type CRSP sequence, is hybridized to a cDNA or other DNA product from a test cell(s). The duplex is treated with a DNA mismatch repair enzyme, and the cleavage products, if any, can be detected from electrophoresis protocols or the like. See, for example, U.S. Patent No. 5,459,039.

In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in CRSP genes. For example, single strand conformation polymorphism (SSCP) may be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids (Orita *et al.* (1989) *Proc Natl. Acad. Sci USA*: 86:2766, see also Cotton (1993) *Mutat Res* 285:125-144; and Hayashi (1992) *Genet Anal Tech Appl* 9:73-79). Single-stranded DNA fragments of sample and control CRSP nucleic acids will be denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence, the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay

may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In a preferred embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility (Keen *et al.* (1991) *Trends Genet* 7:5).

5 In yet another embodiment the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE) (Myers *et al.* (1985) *Nature* 313:495). When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a GC clamp of approximately 40 bp of  
10 high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA (Rosenbaum and Reissner (1987) *Biophys Chem* 265:12753).

Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective  
15 primer extension. For example, oligonucleotide primers may be prepared in which the known mutation is placed centrally and then hybridized to target DNA under conditions which permit hybridization only if a perfect match is found (Saiki *et al.* (1986) *Nature* 324:163); Saiki *et al.* (1989) *Proc. Natl Acad. Sci USA* 86:6230). Such allele specific oligonucleotides are hybridized to PCR amplified target DNA or a number of different  
20 mutations when the oligonucleotides are attached to the hybridizing membrane and hybridized with labeled target DNA.

Alternatively, allele specific amplification technology which depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification may carry the mutation of  
25 interest in the center of the molecule (so that amplification depends on differential hybridization) (Gibbs *et al.* (1989) *Nucleic Acids Res.* 17:2437-2448) or at the extreme 3' end of one primer where, under appropriate conditions, mismatch can prevent, or reduce polymerase extension (Prossner (1993) *Tibtech* 11:238). In addition it may be desirable to introduce a novel restriction site in the region of the mutation to create  
30 cleavage-based detection (Gasparini *et al.* (1992) *Mol. Cell Probes* 6:1). It is anticipated that in certain embodiments amplification may also be performed using Taq ligase for amplification (Barany (1991) *Proc. Natl. Acad. Sci USA* 88:189). In such cases, ligation will occur only if there is a perfect match at the 3' end of the 5' sequence making it possible to detect the presence of a known mutation at a specific site by looking for the  
35 presence or absence of amplification.

The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which may be conveniently used, *e.g.*, in clinical settings to diagnose

patients exhibiting symptoms or family history of a disease or illness involving a CRSP gene.

Furthermore, any cell type or tissue in which CRSP is expressed may be utilized in the prognostic assays described herein.

5

### 3. Monitoring of Effects During Clinical Trials

Monitoring the influence of agents (e.g., drugs, compounds) on the expression or activity of CRSP (e.g., modulation of cellular signal transduction, regulation of gene transcription in a cell involved in development or differentiation, regulation of cellular proliferation) can be applied not only in basic drug screening, but also in clinical trials. For example, the effectiveness of an agent determined by a screening assay as described herein to increase CRSP gene expression, protein levels, or upregulate CRSP activity, can be monitored in clinical trials of subjects exhibiting decreased CRSP gene expression, protein levels, or downregulated CRSP activity. Alternatively, the effectiveness of an agent determined by a screening assay to decrease CRSP gene expression, protein levels, or downregulate CRSP activity, can be monitored in clinical trials of subjects exhibiting increased CRSP gene expression, protein levels, or upregulated CRSP activity. In such clinical trials, the expression or activity of CRSP and, preferably, other genes that have been implicated in, for example, a proliferative disorder can be used as a "read out" or markers of the phenotype of a particular cell.

For example, and not by way of limitation, genes, including CRSP, that are modulated in cells by treatment with an agent (e.g., compound, drug or small molecule) which modulates CRSP activity (e.g., identified in a screening assay as described herein) can be identified. Thus, to study the effect of agents on proliferative disorders, developmental or differentiative disorder, or hematopoietic disorder, for example, in a clinical trial, cells can be isolated and RNA prepared and analyzed for the levels of expression of CRSP and other genes implicated in the proliferative disorder, developmental or differentiative disorder, or hematopoietic disorder, respectively. The levels of gene expression (i.e., a gene expression pattern) can be quantified by Northern blot analysis or RT-PCR, as described herein, or alternatively by measuring the amount of protein produced, by one of the methods as described herein, or by measuring the levels of activity of CRSP or other genes. In this way, the gene expression pattern can serve as a marker, indicative of the physiological response of the cells to the agent. Accordingly, this response state may be determined before, and at various points during treatment of the individual with the agent.

In a preferred embodiment, the present invention provides a method for monitoring the effectiveness of treatment of a subject with an agent (e.g., an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug



candidate identified by the screening assays described herein) comprising the steps of (i) obtaining a pre-administration sample from a subject prior to administration of the agent; (ii) detecting the level of expression of a CRSP protein, mRNA, or genomic DNA in the preadministration sample; (iii) obtaining one or more post-administration samples from the subject; (iv) detecting the level of expression or activity of the CRSP protein, mRNA, or genomic DNA in the post-administration samples; (v) comparing the level of expression or activity of the CRSP protein, mRNA, or genomic DNA in the pre-administration sample with the CRSP protein, mRNA, or genomic DNA in the post administration sample or samples; and (vi) altering the administration of the agent to the subject accordingly. For example, increased administration of the agent may be desirable to increase the expression or activity of CRSP to higher levels than detected, i.e., to increase the effectiveness of the agent. Alternatively, decreased administration of the agent may be desirable to decrease expression or activity of CRSP to lower levels than detected, i.e. to decrease the effectiveness of the agent. According to such an embodiment, CRSP expression or activity may be used as an indicator of the effectiveness of an agent, even in the absence of an observable phenotypic response.

#### C. Methods of Treatment:

The present invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant CRSP expression or activity. With regards to both prophylactic and therapeutic methods of treatment, such treatments may be specifically tailored or modified, based on knowledge obtained from the field of pharmacogenomics. "Pharmacogenomics", as used herein, refers to the application of genomics technologies such as gene sequencing, statistical genetics, and gene expression analysis to drugs in clinical development and on the market. More specifically, the term refers the study of how a patient's genes determine his or her response to a drug (e.g., a patient's "drug response phenotype", or "drug response genotype".) Thus, another aspect of the invention provides methods for tailoring an individual's prophylactic or therapeutic treatment with either the CRSP molecules of the present invention or CRSP modulators according to that individual's drug response genotype. Pharmacogenomics allows a clinician or physician to target prophylactic or therapeutic treatments to patients who will most benefit from the treatment and to avoid treatment of patients who will experience toxic drug-related side effects.

### 1. Prophylactic Methods

In one aspect, the invention provides a method for preventing in a subject, a disease or condition associated with an aberrant CRSP expression or activity, by administering to the subject an agent which modulates CRSP expression or at least one CRSP activity. Subjects at risk for a disease which is caused or contributed to by aberrant CRSP expression or activity can be identified by, for example, any or a combination of diagnostic or prognostic assays as described herein. Administration of a prophylactic agent can occur prior to the manifestation of symptoms characteristic of the CRSP aberrancy, such that a disease or disorder is prevented or, alternatively, delayed in its progression. Depending on the type of CRSP aberrancy, for example, an CRSP agonist or CRSP antagonist agent can be used for treating the subject. The appropriate agent can be determined based on screening assays described herein. The prophylactic methods of the present invention are further discussed in the following subsections.

### 2. Therapeutic Methods

Another aspect of the invention pertains to methods of modulating CRSP expression or activity for therapeutic purposes. The modulatory method of the invention involves contacting a cell with an agent that modulates one or more of the activities of CRSP protein activity associated with the cell. An agent that modulates CRSP protein activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring target molecule of a CRSP protein, a peptide, a CRSP peptidomimetic, or other small molecule. In one embodiment, the agent stimulates one or more CRSP protein activity. Examples of such stimulatory agents include active CRSP protein and a nucleic acid molecule encoding CRSP that has been introduced into the cell. In another embodiment, the agent inhibits one or more CRSP protein activity. Examples of such inhibitory agents include antisense CRSP nucleic acid molecules and anti-CRSP antibodies. These modulatory methods can be performed *in vitro* (e.g., by culturing the cell with the agent) or, alternatively, *in vivo* (e.g., by administering the agent to a subject). As such, the present invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant expression or activity of a CRSP protein or nucleic acid molecule. In one embodiment, the method involves administering an agent (e.g., an agent identified by a screening assay described herein), or combination of agents that modulates (e.g., upregulates or downregulates) CRSP expression or activity. In another embodiment, the method involves administering a CRSP protein or nucleic acid molecule as therapy to compensate for reduced or aberrant CRSP expression or activity.

Stimulation of CRSP activity is desirable in situations in which CRSP is abnormally downregulated and/or in which increased CRSP activity is likely to have a

beneficial effect. Likewise, inhibition of CRSP activity is desirable in situations in which CRSP is abnormally upregulated and/or in which decreased CRSP activity is likely to have a beneficial effect. One example of such a situation is where a subject has a disorder characterized by aberrant development or cellular differentiation. Another example of such a situation is where the subject has a proliferative disease (e.g., cancer) or a disorder characterized by an aberrant hematopoietic response. Yet another example of such a situation is where it is <sup>desirable</sup> ~~desireable~~ to <sup>achieve</sup> ~~generate~~ tissue regeneration in a subject (e.g., where a subject has undergone brain or spinal cord injury and it is desirable to regenerate neuronal tissue in a regulated manner.)

10

### 3. Pharmacogenomics

The CRSP molecules of the present invention, as well as agents, or modulators which have a stimulatory or inhibitory effect on CRSP activity (e.g., CRSP gene expression) as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically) disorders (e.g, proliferative or developmental disorders) associated with aberrant CRSP activity. In conjunction with such treatment, pharmacogenomics (i.e., the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) may be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, a physician or clinician may consider applying knowledge obtained in relevant pharmacogenomics studies in determining whether to administer an CRSP molecule or CRSP modulator as well as tailoring the dosage and/or therapeutic regimen of treatment with an CRSP molecule or CRSP modulator.

Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See e.g., Eichelbaum, M., *Clin Exp Pharmacol Physiol*, 1996, 23(10-11) :983-985 and Linder, M.W., *Clin Chem*, 1997, 43(2):254-266. In general, two types of pharmacogenetic conditions can be differentiated. Genetic conditions transmitted as a single factor altering the way drugs act on the body (altered drug action) or genetic conditions transmitted as single factors altering the way the body acts on drugs (altered drug metabolism). These pharmacogenetic conditions can occur either as rare genetic defects or as naturally-occurring polymorphisms. For example, glucose-6-phosphate dehydrogenase deficiency (G6PD) is a common inherited enzymopathy in which the main clinical complication is haemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

35

One pharmacogenomics approach to identifying genes that predict drug response, known as "a genome-wide association", relies primarily on a high-resolution

map of the human genome consisting of already known gene-related markers (e.g., a "bi-allelic" gene marker map which consists of 60,000-100,000 polymorphic or variable sites on the human genome, each of which has two variants.) Such a high-resolution genetic map can be compared to a map of the genome of each of a statistically  
5 significant number of patients taking part in a Phase II/III drug trial to identify markers associated with a particular observed drug response or side effect. Alternatively, such a high resolution map can be generated from a combination of some ten-million known single nucleotide polymorphisms (SNPs) in the human genome. As used herein, a "SNP" is a common alteration that occurs in a single nucleotide base in a stretch of  
10 DNA. For example, a SNP may occur once per every 1000 bases of DNA. A SNP may be involved in a disease process, however, the vast majority may not be disease-associated. Given a genetic map based on the occurrence of such SNPs, individuals can be grouped into genetic categories depending on a particular pattern of SNPs in their individual genome. In such a manner, treatment regimens can be tailored to groups of  
15 genetically similar individuals, taking into account traits that may be common among such genetically similar individuals.

Alternatively, a method termed the "candidate gene approach", can be utilized to identify genes that predict drug response. According to this method, if a gene that encodes a drug's target is known (e.g., a CRSP protein or CRSP receptor of the present  
20 invention), all common variants of that gene can be fairly easily identified in the population and it can be determined if having one version of the gene versus another is associated with a particular drug response.

As an illustrative embodiment, the activity of drug metabolizing enzymes is a major determinant of both the intensity and duration of drug action. The discovery of  
25 genetic polymorphisms of drug metabolizing enzymes (e.g., N-acetyltransferase 2 (NAT 2) and cytochrome P450 enzymes CYP2D6 and CYP2C19) has provided an explanation as to why some patients do not obtain the expected drug effects or show exaggerated drug response and serious toxicity after taking the standard and safe dose of a drug. These polymorphisms are expressed in two phenotypes in the population, the extensive  
30 metabolizer (EM) and poor metabolizer (PM). The prevalence of PM is different among different populations. For example, the gene coding for CYP2D6 is highly polymorphic and several mutations have been identified in PM, which all lead to the absence of functional CYP2D6. Poor metabolizers of CYP2D6 and CYP2C19 quite frequently experience exaggerated drug response and side effects when they receive standard doses.  
35 If a metabolite is the active therapeutic moiety, PM show no therapeutic response, as demonstrated for the analgesic effect of codeine mediated by its CYP2D6-formed metabolite morphine. The other extreme are the so called ultra-rapid metabolizers who

do not respond to standard doses. Recently, the molecular basis of ultra-rapid metabolism has been identified to be due to CYP2D6 gene amplification.

Alternatively, a method termed the "gene expression profiling", can be utilized to identify genes that predict drug response. For example, the gene expression of an animal dosed with a drug (e.g., a CRSP molecule or CRSP modulator of the present invention) can give an indication whether gene pathways related to toxicity have been turned on.

Information generated from more than one of the above pharmacogenomics approaches can be used to determine appropriate dosage and treatment regimens for prophylactic or therapeutic treatment an individual. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with a CRSP molecule or CRSP modulator, such as a modulator identified by one of the exemplary screening assays described herein.

This invention is further illustrated by the following examples which should not be construed as limiting. The contents of all references, patents and published patent applications cited throughout this application are hereby incorporated by reference.

20

## EXAMPLES

### **Example 1: Isolation And Characterization of Human CRSP-1 cDNA**

In this example, the isolation and characterization of the gene encoding human CRSP-1 (also referred to as "CRISPY-1" or "TANGO 59") is described.

#### **Isolation of the human CRSP-1 cDNA**

The invention is based at least in part on the discovery of a human gene encoding a secreted protein, referred to herein as Cysteine Rich Secreted Protein-1 (CRSP-1). A partial cDNA was isolated using a Signal Sequence Trap method. This methodology takes advantage of the fact that molecules such as CRSP have an amino terminal signal sequence which directs certain secreted and membrane-bound proteins through the cellular secretory apparatus.

Briefly, a randomly primed cDNA library using mRNA prepared from human fetal brain tissue (Clontech, Palo Alto CA) was made by using the Stratagene-ZAP-cDNA Synthesis™ kit, (catalog #20041). The cDNA was ligated into the mammalian expression vector pTrap adjacent to a cDNA encoding placental alkaline phosphatase lacking a secretory signal. The plasmids were transformed into *E. coli* and DNA was

prepared using the Wizard™ DNA purification kit (Promega). DNA was transfected into COS-7 cells with lipofectamine™ (Gibco-BRL). After 48 hours incubation the COS cell supernatants were assayed for alkaline phosphatase on a Wallac Micro-Beta scintillation counter using the Phospha-Light™ kit (Tropix Inc. Catalog #BP300). The individual plasmid DNAs scoring positive in the COS cell Alkaline Phosphatase secretion assay were further analyzed by DNA sequencing using standard procedures.

Using a partial cDNA isolated by the above-described method, a full length cDNA encoding human CRSP-1 was cloned. The nucleotide sequence encoding the full length human CRSP-1 protein is shown in Figure 1 and is set forth as SEQ ID NO: 1. The full length protein encoded by this nucleic acid is comprised of about 350 amino acids and has the amino acid sequence shown in Figure 1 and set forth as SEQ ID NO:2. The coding portion (open reading frame) of SEQ ID NO:1 is set forth as SEQ ID NO:3.

#### Analysis of Human CRSP-1

Determination of the hydrophobicity profile of human CRSP-1 having the amino acid sequence set forth in SEQ ID NO:2 indicated the presence of a hydrophobic region from about amino acid 1 to about amino acid 23 of SEQ ID NO:2. The presence of the signal sequence, in addition to the fact that CRSP-1 has been identified using a Signal Sequence Trap system, indicates that CRSP-1 is a secreted protein. Furthermore, the prediction of such a signal peptide and signal peptide cleavage site can be made, for example, utilizing the computer algorithm SIGNALP (Henrik, et al. (1997) *Protein Engineering* 10:1-6).

Furthermore, when the cDNA encoding human CRSP-1 was expressed in bacterial cells with a C-terminal FLAG tag, the FLAG-tagged CRSP-1 protein product was found to be secreted in to the cellular medium.

Examination of the cDNA sequence depicted in Figure 1 shows that human CRSP-1 is particularly rich in cysteine residues. As shown in Figure 1, CRSP-1 contains 20 cysteine residues located between amino acid 147 and amino acid 284 of SEQ ID NO: 2. These cysteine residues can possibly form 10 disulfide bridges.

A BLAST search (Altschul et al. (1990) *J. Mol. Biol.* 215:403) of the nucleotide and the amino acid sequences of CRSP-1 has revealed that CRSP-1 is significantly similar to a chicken cDNA encoding a protein of unknown function having GenBank Accession No. D26311. This cDNA was isolated from a chicken lens cDNA library and was shown to be expressed in lens fibers and lens epithelium, but not in neural retina nor in liver cells. (Sawada et al. (1996) *Int. J. Dev. Biol.* 40:531). CRSP-1 and the chicken protein have 56% amino acid sequence identity and 72% amino acid sequence similarity. The amino acid sequence similarity between the chicken protein and human CRSP-1 is particularly high in the cysteine-rich domain of CRSP-1 which is located between amino

acids 147 and 284 of SEQ ID NO: 2. In particular, the 20 cysteine residues of CRSP-1 located in this region are present in the chicken protein (see Figure 3).

#### Tissue Distribution of CRSP-1 mRNA

5        This Example describes the tissue distribution of CRSP-1 mRNA, as determined by Northern blot hybridization.

Northern blot hybridizations with the various RNA samples were performed under standard conditions and washed under stringent conditions, i.e., 0.2xSSC at 65°C. In each sample, the probe hybridized to a single RNA of about 2.5 kb. The results of  
10    hybridization of the probe to various mRNA samples are described below.

Hybridization of a Clontech Fetal Multiple Tissue Northern (MTN) blot (Clontech, LaJolla, CA) containing RNA from fetal brain, lung, liver, and kidney indicated the presence of high levels of CRSP-1 mRNA in fetal brain, lung, and slightly lower levels of CRSP-1 mRNA in fetal kidney. However, no significant level of CRSP-  
15    1 mRNA was found in fetal liver.

Hybridization of a Clontech human Multiple Tissue Northern (MTN) blot (Clontech, LaJolla, CA) containing RNA from adult heart, brain, placenta, lung, liver, skeletal muscle, kidney, and pancreas with a human CRSP-1 probe indicated the presence of high levels of CRSP-1 mRNA in heart, slightly lower levels in brain, and  
20    much lower levels in placenta and lung. Some CRSP-1 mRNA was also found in adult skeletal muscle. However, no significant levels of CRSP-1 mRNA was observed in adult liver, kidney, or pancreas. Interestingly, the chicken gene which is homologous to CRSP-1 was not expressed at detectable levels in liver either (Sawada et al. (1996) Int. J. Dev. Biol. 40:531).

25        Further hybridization of a Clontech human Multiple Tissue Northern (MTN) blot (Clontech, LaJolla, CA) including RNA from adult spinal cord revealed high levels of expression of CRSP-1 in mRNA isolated from adult spinal cord.

In situ analysis further revealed the following expression patterns when tissue sections were hybridized with CRSP-1 probes. CRSP-1 mRNA was expressed in heart, lung, aorta, eye (retina and lens), and brain (cortex) in 14.5 day mouse embryos. CRSP-  
30    1 mRNA was expressed in heart, lung, eye (retina and lens), brain (cortical), and cartilage (foot, trachea, larynx, head, sternum, and spinal column) in 1.5 day post natal mouse tissues. CRSP-1 mRNA was expressed in brain (cortex, hippocampus, brainstem), heart (atria only), eye (retina and lens outer layer) in adult mouse tissues.

35        Thus, CRSP-1 is expressed in a tissue specific manner, with the strongest expression observed in brain, heart, and spinal cord.

### CRSP-1 protein expression data

#### **Constructs**

Expression constructs for two forms of CRSP-1 were prepared using the mammalian expression vector pMET-stop. Form-1 comprised a cDNA incorporating the complete 350aa CRSP-1 protein coding sequence (CRSP-1flag.long) and form-2 comprised the entire CRSP-1 protein coding sequence except for the final 18 amino acids (CRSP-1flag.short). A C-terminal sequence encoding the FLAG epitope (DYKDDDDK) was added to both CRSP-1 forms for ease of detection and purification. CRSP-1flag cDNAs were generated by PCR from a full length CRSP-1 cDNA template and ligated into pMET-stop using EcoR1 and Sal1 restriction sites.

#### **Trial transfection - small scale expression**

Expression constructs for CRSP-1flag.long and CRSP-1flag.short were transfected into 293T cells using 10 µl of lipofectamine (GIBCO/BRL) and 2 µg of DNA per well of a 6-well plate of cells which were 70-80% confluent. After 5 hours at 37°C, cells were fed with 1ml of 20%FCS/DMEM. After incubation overnight at 37°C, cells were conditioned in 1ml OptiMEM for 48 hours at 37°C. Samples of supernatant and cell pellets were solubilized in boiling SDS-PAGE gel buffer, run out on a 4-20% SDS-PAGE gel, transferred to a nylon membrane and probed with the anti-FLAG monoclonal antibody M2. Samples from both supernatant and pellet samples showed significant immunoreactivity within a molecular weight range of 40-65 kDa on autoradiographic film using a HRP conjugated secondary antibody and ECL detection reagents. Thus, both forms of CRSP-1 tested are secreted from 293T cells thereby confirming experimentally that CRSP-1 is a secreted protein. It should be noted that the molecular weights of both forms of CRSP-1 tested are greater than predicted from the amino acid sequence, suggesting that the CRSP-1 proteins secreted by 293T cells may be glycosylated. This is consistent with the presence of four potential sites for N-linked glycosylation in the CRSP-1 protein.

#### **Large scale CRSP-1 protein production**

For scale-up of CRSP-1flag.long protein expression, 30 x 150mm plates of 293T cells at 70-80% confluence were transfected with 27 µg DNA, 100 µl lipofectamine in 18ml OptiMEM for 5 hours at 37°C. 18 ml of 10%FCS/DMEM was added to each plate and incubated overnight at 37°C. 24 hours after the start of transfection, transfection supernatant was aspirated and 35 mls OptiMEM was added to each plate and the plates incubated at 37°C for 72 hours. Conditioned medium was harvested, spun at 4000 rpm for 30 min. at 4°C, and filtered through a 0.45 micron filter unit. 1100 ml was passed over a 1.6 x 10 cm anti-FLAG M2 affinity column pre-equilibrated in PBS pH7.4 buffer



at a flow rate of 2.0 ml per minute. After washing with 200 ml of PBS pH 7.4 buffer, bound material was eluted by a step of 200 mM Glycine pH 3.0 buffer and 0.5 ml fractions collected. Upon elution, a significant protein peak was detected by absorbance at 280nm. Samples corresponding to conditioned medium, flow through and eluted fractions were analyzed by Coomassie blue and silver stained SDS-PAGE and by western blot analysis as described above. Significant immunoreactivity within a molecular weight range of 40-65 kDa was detected in conditioned medium and eluted fractions but not in the flow through sample, indicating that the secreted CRSP-1 flag.long protein bound to the affinity column specifically and was eluted efficiently by the described conditions. Coomassie blue staining of SDS-PAGE gels suggested that the predominant immunoreactive protein constituted >90% of the protein present in the bound and eluted protein peak. Peak fractions of eluted protein were pooled and dialysed against Phosphate Buffered Saline resulting in a 4 ml volume of recombinant CRSP-1 flag.long protein at a concentration of approximately 1mg/ml.

#### **Example 2: Isolation And Characterization of Human CRSP-2, CRSP-3, and CRSP-4 cDNA**

To identify novel proteins related to CRSP-1, the human CRSP-1 amino acid sequence was used to search the private and dbEST databases using TBLASTN (WashUversion, 2.0, BLOSUM62 search matrix). From this search, four distinct partial protein sequences were identified which displayed significant homology to human CRSP-1. These sequences were designated as partial sequences for the novel hCRSPs; hCRSP-2, h-CRSP-3, h-CRSP-4 and h-CRSP-like-n.

hCRSP-2 partial protein sequence was derived from a single dbEST clone with accession number AA565546. This clone was subsequently obtained from the IMAGE collection and sequenced fully to define the entire hCRSP-2 sequence depicted in Figure 2.

hCRSP-3 partial protein sequence was derived from a jthKb075a10 clone with sequence identification code jthKb075a10. This clone was sequenced further and this sequence information combined with additional sequences from dbEST using the assembly program Phrap (P. Green, U. Washington) to define the entire hCRSP-3 sequence depicted in Figure 3. DNA for the Millennium clone jthKb075a10 was deposited with the ATCC as Accession No. 98633 on January 16, 1998.

h-CRSP-4 partial protein sequence was derived from a single dbEST clone with accession number W55979. This clone was subsequently obtained from the IMAGE collection and sequenced fully to define a hCRSP-4 sequence depicted in Figure 4.

INS C3 C35

h-CRSP-like-n partial protein sequence was derived from a single dbEST clone with accession number AA397836. This clone was subsequently obtained from the IMAGE collection and sequenced fully to define the entire hCRSP-like-n sequence depicted in Figure 5.

5

#### Structure of the CRSP Family proteins

An alignment of the amino acid sequences of human CRSP-1, human CRSP-2, human CRSP-3, and human CRSP-4 is shown in Figure 6. Amino acid residues which are conserved between human CRSP family members are boxed. The 20 conserved cysteine residues are indicated by an asterix. The cysteine-rich domains are indicated as CRD-1 and CRD-2. The domain structure of the full length CRSP proteins of the present invention are depicted in Figure 7.

#### 15 **Equivalents**

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

20